CRWN FAMILY PROTEINS REGULATE NUCLEAR ORGANIZATION AND NUCLEAR FUNCTION IN ARABIDOPSIS THALIANA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

Haiyi Wang

May 2014

© 2014 Haiyi Wang

CRWN FAMILY PROTEINS REGULATE NUCLEAR ORGANIZATION AND NUCLEAR FUNCTION IN ARABIDOPSIS THALIANA

Haiyi Wang Ph. D.

Cornell University [2014]

Abstract:

The molecular components and processes that shape and organize nuclei in plant cells are poorly understood. This thesis describes genetic, cytological, and biochemical studies of CRWN (CROWDED NUCLEI) proteins, which are required for proper nuclear structure in the flowering plant, *Arabidopsis thaliana*. These plant-specific proteins feature a long coiled-coil motif, with a conserved C-terminal domain. CRWN proteins are expressed primarily in proliferating tissues, and are located at the nuclear periphery. CRWN1 and CRWN4 belong to a nuclear fraction resistant to high salt and mild detergent extraction. I hypothesize that CRWN proteins are structural components of plant nuclei.

Genetic analysis of the whole family of *crwn* mutants in *Arabidopsis thaliana* revealed a variety of phenotypic changes, including altered nuclear shape, reduced nuclear size,

and heterochromatin aggregation or dispersion, mildly decreased endopolyploidy levels, and increased nuclear DNA density. In addition, some *crwn* mutants were dwarfed with early flowering times, shorter internodes, heavier branches, and delayed senescence. A subsequent mRNA-seq profiling in representative *crwn* mutants illustrated genome-wide transcriptional mis-regulation, and identified candidate genes responsible for phenotypic changes in *crwn* mutants. I propose that the loss of CRWN proteins primarily alters nuclear organization, leading to alterations in gene expression.

Phylogenetic analysis partitioned this protein family into *CRWN1*-like and *CRWN4*-like sub-categories, and non-redundant morphological alterations observed in *crwn1* versus *crwn4* mutants further support this idea. Nonetheless, the transcriptomic data uncovered shared profiles of mis-expressed loci regulated by *CRWN1*-like and *CRWN4*-like genes. Moreover, regulatory relationships among CRWN paralogs exist on both the mRNA and protein level. A physical interaction between CRWN1 and CRWN4 proteins was demonstrated by immunoprecipitation experiments. These findings suggest that different CRWN proteins function together to maintain nuclear organization in plants.

BIOGRAPHICAL SKETCH

Haiyi Wang earned the bachelor degree of biological sciences from Fudan University in 2006. In college, Haiyi studied the ecology and conservation of grassland in Inner Mongolia, assisted curating the herbarium in Fudan University, and used molecular ecology tools to explore the genetic diversity in wild soybean populations from Yunnan Provinces.

Later, she joined the Graduate Program of Plant Biology in Cornell University, and started her thesis project in Richards' lab since Jan 2009, focusing on nuclear organization and function in *Arabidopsis thaliana*. In graduate school, Haiyi explored various aspects of epigenetics, nuclear organization and evolution in *Arabidopsis thaliana*, and published first author article "Arabidopsis CROWDED NUCLEI (CRWN) proteins are required for nuclear size control and heterochromatin organization" on BMC Plant Biology. She attended many symposia and conferences, and presented posters on Chromosome Dynamics Gordon Conference (2010) and Plant Epigenetics, Stress and Evolution CSHA (Cold Spring Harbor Asia) Conference (2012). She had assisted teaching both graduate and undergraduate level courses, and supervised undergraduate students in the lab on CRWN projects.

Haiyi also actively participates in academic services, including volunteering in organization of conferences, serving on Graduate School panels, Seminar Committees, in Boyce Thompson Institute, and Plant Biology Graduate Student Association.

ACKNOWLEDGMENTS

It has been unusual seven years in my life. I first want to express my gratefulness to Dr. Eric Richards, a very special scientist, for guiding me into the fun side of biology. I can never forget the moments when a mystic smile fulfilled with satisfaction flashes through his eyes, when the curiosity is lighted up by the beauty of nature.

I cherish the efforts my parents had made recent years as a remote support.

I want to say thank you to Dr. Wojtech Pawlowski, Jun Liu, June Nasrallah for their guidance and equipment supports on my research; to Dr. Jeff Doyle, Tom Owens and Jian Hua for their advices along my graduate life; to people in Dr. Tom Brutnell and Dr. Fei Zhangjun lab for many helpful discussions; to my current lab members as well as Travis Dittmer, a previous graduate student who initiated the CRWN project, for the blended intellectual inputs to my studies. My experiments could not be completed without the help from BTI and Cornell colleagues, including Lauren Dedow, Silin Zhong, Choonlin Tang, Patrick Boyle, Mamta Srivastava, and I got advices for the computational analysis from Dr. Qi Sun (CBSU) and bioinformaticians in BTI.

Life in these years has been enriched with many adventures. I would love to express my thankfulness to Molly Shook, Blake Meyers, Caroline Sartain and Aziana Ismail, for breaking the ice of diverse cultures with their genuineness, respects, and helps. I highly appreciated Chongshan Huang, Xiao Chen, Youyu Zhang, Xiangyi Xu,

Lingyan Jin, Nan Zheng and Ting Shi for their company, regardless where they are in the world; and Lin Zhuang, Shu Huang and Sheldon Zhang, for sharing their immigration experiences with me. I will also remember the company and support from my Cornell fellows Xiaohua Yang, Xinsying Jiang, Jing Jin, Jun Qian, Lin Xu and Yuan Si.

My stipend, tuition and research were funded by Olin Fellowship, TAship, NSF grants and BTI start-up fund for the Richards' Lab.

TABLE OF CONTENTS

LIST OF FIGURES

Chapter 1

Chapter 2

Chapter 4

LIST OF TABLE

CHAPTER ONE

INTRODUCTION: THE NUCLEAR ORGANIZATION AND THE REGULATION OF NUCLEAR FUNCTION

History

The first description of the cell nucleus as an independent organelle was recorded in 1682 when Antonie van Leeuwenhoek drew a "lumen" structure that he observed in cod and salmon erythrocytes [1]. In 1802, Franz Bauer also described the putative cell nucleus in orchid, *Bletia tankervilliae* [2], and this observation was later popularized by the botanist Robert Brown when he presented to the Linnaean Society of London a structure that he called the "areola," an opaque area in the outer layer cells of flowers [3].

These discoveries were followed by many studies and continuing debate on the functional role of the nucleus. In the year of 1838, Matthias Schleiden suggested that the nucleus was a structure he termed the "cytoblast," or cell builder, that gave rise to cells to form the next cell generation. As a strong opponent of the cytoblast view, Franz Meyen proposed that cells multiply via cell division, and hypothesized that many cells would not have nuclei. In 1850s Robert Remak and Rudolf Virchow argued against the *de novo* cytoblast hypothesis and put forward a new paradigm that new cells can only originate from the division of pre-existing cells [4].

An important transition in cell theory occurred in the 1870s, when Oscar Hertwig showed that the sperm nucleus entered the oocyte and fused with its nucleus during sea urchin egg fertilization. Similar observations were later made in plants by Eduard Strasburger, and this work laid the foundation for the discovery of the essential function that the nucleus plays in heredity. In the beginning of the 20th century, Mendelian rules and the chromosome theory of heredity were developed, and the nucleus was recognized as the primary carrier of genetic information [4]. From that point forward attention and efforts were focused on understanding the structure and function of genes and how they control the phenotype of the cell and the organism.

However, in the past few decades, research on nuclear organization has become popular again, uncovering new levels of regulation of transcription and other fundamental nuclear processes. In this introductory chapter, I will first selectively review work on the structural composition and functional impact of major nuclear components in non-plant systems, with a emphasis on three different topics: 1) the principles of chromosome and chromatin organization; 2) the molecular components of the nuclear envelope and nuclear pores; and 3) the different constituents of the nucleoskeleton. Then, I will turn my attention to work in plants, primarily Arabidopsis, which is directly relevant to my research.

Review of nuclear organization in non-plant systems

Chromosomal organization is important for transcriptional regulation

Walther Flemming, who first described the process of mitosis in the mid-19th century, is credited with the discovery of chromosomes as visible or "stainable" units in the cell [5]. Carl Rabl later described a distinct bouquet-like organization of chromosomes that is often seen after cell division, with centromeres clustered on one side of the nucleus and chromosome ends located on the opposite side [6]. This observation indicated that the hereditary material had some higher-order three-dimensional organization. Subsequent studies uncovered substructures within chromosomes. Cytological analysis of interphase nuclei in the liverwort *Pellia epiphylla* led Emil Heitz to identify two types of chromatin - the invisible or poorly stained regions of chromosomes as "euchromatin," and the deeply stained visible knots as "heterochromatin" [7]. In most eukaryotic cells, decondensed euchromatin is associated with active transcription, and resides in the interior of the nucleus; while the tightly packed heterochromatin is associated with transcriptional repression, and is usually localized to the nuclear periphery [8]. However, exceptions exist. For example, in retinal rod cells of mice, the typical pattern is reversed, and heterochromatin is located in the interior of the nucleus. This unusual arrangement is thought to maximize the transparency of the rod cells to enhance the light transmission to the photoreceptors in these nocturnal mammals that need to see in dim light [8].

The mechanisms that mediate the spatial compartmentalization of eu- and hetero-chromatin are not well understood, but some clues have been generated from work in Drosophila and human demonstrating that an interaction between Heterochromatin Protein 1 (HP1) and the lamin B receptor help anchor heterochromatin to the nuclear envelope [9-11]. This non-random positioning of chromatin also impacts transcription [12, 13]. One classic example is position-effect variegation (PEV) of the expression of genes after relocation to chromosomal positions close to heterochromatin. Hermann Müller discovered PEV from the study of X-ray-induced chromosomal rearrangements in *Drosophila* [14]. He observed that the function of the euchromatic gene *white*, which was now placed adjacent to heterochromatin from the centromeric region due to the rearrangement, became variably inactivated. This inactivation led to clonal sectors of differential pigment in the insect's compound eye. Later work demonstrated that the reversible inactivation of the white gene was due to heterochromatin spreading across the euchromatin/heterochromatin breakpoint, silencing *white* and giving rise to a variegated phenotype [15]. Insertion of transposable elements into euchromatic regions can also promote the formation of local heterochromatin and result in PEV [16]. Some researchers propose that selection has acted on these types of chromosomal position effects and resulted in the linear grouping of co-regulated genes along chromosome arms. One particularly striking example is the *HOX* genes that are ordered along the chromosome corresponding to their spatio-temporal expression for organ development [17]. Genomic studies in higher eukaryotes have illustrated the occurrence of clusters of highly-transcribed genes or tissue-specific genes along the chromosome, suggesting a position effect on transcriptional regulation [18].

The non-random compartmentalization of different regions along the chromosomes is likely achieved via alternate layers of compaction, which can also control or at least affect gene expression. A view emerged in the 1980s that chromatin fibers were organized into loop domains, anchored via S/MAR (scaffold or matrix attachment regions) sequences to a scaffolding system in interphase nuclei [19]. These sequences were first identified by purifying DNA fragments that remain associated with the "nuclear matrix," generally defined as a salt- or detergent-resistant remnant left after extraction of nuclei [19]. S/MARs are typically a few hundred base pairs in length, enriched with AT nucleotides, and localized in the noncoding regions flanking genes [20]. Nonetheless, some studies have reported that S/MAR regions are transcriptionally active [21, 22]. The hypothesis that higher-order chromatin structure consists of confined and discrete chromosome territories is supported by the development of Fluorescence *in situ* Hybridization (FISH) techniques coupled with "chromosome painting" [23]. The more recent development of chromosome conformation capture techniques reveals that specific long-range interactions can occur, but that local interactions predominate. These Hi-C and related techniques have generally supported the looping model [24]. Simulations of the Hi-C data suggest that chromatin is organized like a polymer arranged in a linear array of loops compressed longitudinally [25]. These types of studies have the power to identify genes or regulatory sequences widely separated in *cis*, if they co-localize to a common niche in the nuclei for transcriptional regulation. One such example exists in erythroid cells in mouse, where the transcription factor Klf1 co-associates with genes that the protein

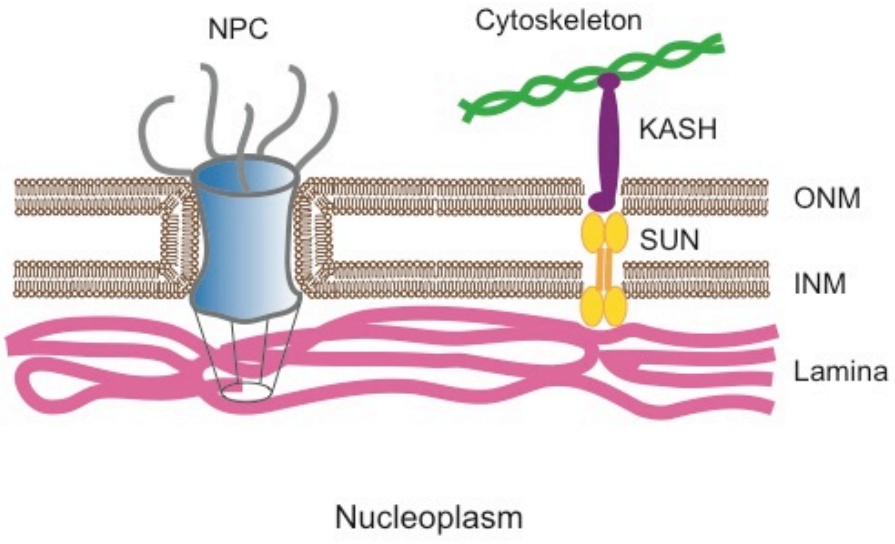
regulates at specialized transcription factories [26]. In parallel with the development of new genomics approaches to identify sequences involved in higher-order chromatin organization, computational tools are being applied to the problem. For example, algorithms were used to identify matrix attachment regions (MARS) with an effect on gene activation in the mammalian genome, and these MARS are being engineered into constructs for generating transgenic mice in an effort to alleviate the epigenetic silencing of transgenes and boost the expression of recombinant proteins [27]. The human ENCODE project surveyed all suspected chromatin boundary regions, and discovered 453 nuclear scaffold attachment sites predominately near expressed genes [28]. This highly compartmentalized and cell type-specific chromosome organization appears to be limited to interphase. In metaphase, a locus-independent homogenous folding state exists among all chromosomes in various cell types [24].

General physical mechanisms may also affect the positioning of sub-nuclear compartments, and consequentially affect biological functions [29]. It has been shown that macromolecular crowding contributes to the organization of nucleoli and promyelocytic leukemia (PML) nuclear bodies [30]. Also, osmotic perturbations can alter the local compaction state of chromatin and affect the formation and maintenance of heterochromatin packaging [30]. However, conformational change based on biophysical properties is insufficient to initiate or maintain the densely packed heterochromatin structures. In this case, specific regulatory proteins are still required for regulating chromatin states independent of organizing principles imposed by molecular crowding [30].

The nuclear envelope regulates chromosome positioning and transcription

The nuclear envelope defines the nucleus by its double bilayers of lipid molecules, with embedded nuclear pores for macromolecular trafficking (Figure 1.1). The outer nuclear membrane (ONM) resembles and connects to the endoplasmic reticulum, and the inner nuclear membrane (INM) is involved in dynamic nuclear functions. As noted earlier in this chapter, transcriptionally silent heterochromatin typically localizes close to the nuclear envelope or the nucleolus, but is excluded from the area adjacent to nuclear pores. Examples of heterochromatin preferentially localized to the nuclear periphery in mammals and yeast include repetitive centromeric sequences and transcriptionally silent telomeric chromatin [31], while highly transcribed genes are commonly found at nuclear pores [32].

Various anchoring mechanisms and pathways participate in position-dependent transcriptional regulation in the nucleus. In yeast, the SUN-domain protein, Mps3, anchors silent telomere chromatin to the nuclear rim. In *mps3* mutants, telomeres become partially detached from the periphery and epigenetic repression of the telomeric regions is abrogated [33]. In a similar manner, the yeast LEM (LAP2, emerin, MAN1) protein Src1 (homolog of Man1 in mammals) is co-localized with sub-telomeric chromatin and *src1* mutants mis-regulate the expression of subtelomeric genes [34]. In Drosophila and human cells, lamin (see below) proteins are important for positioning genomic regions at the nuclear periphery, and genome-wide profiling methods using tagging techniques, such as DamID, have revealed an enrichment of



Adapted from Goldman lab website (NPC, lamin, nuclear envelope) and B Burke *et al.* Nature Review MCB 2013

NPC: Nuclear Pore Complex
KASH: Klarsicht, ANC-1, Syne Homology domain containing proteins
SUN: SUN-domain containing proteins
ONM: Outer Nuclear Membrane
INM: Innter Nuclear Membrane

Figure 1.1 Organization of the nuclear envelope in animal cells

Figure 1.1 This figure describes the important components of the nuclear envelope in animals. Double lipid layers separate the nucleoplasm from the cytoplasm. Nuclear Pore Complexes (NPC) are imbedded in the nuclear envelope, with the nuclear baskets of the pore complex being oriented toward the inner nuclear membrane (INM) and the cytoplasmic filaments positioned at the cytoplasmic face with the the outer nuclear membrane (ONM). The SUN and KASH domain proteins form the LINC complexes (Linker of Nulceoskeleton and Cytoskeleton), bridging the nuclear lamina and cytoskeleton across the nuclear envelope.

transcriptionally silent loci at lamin-associated domains (LADs) [35, 36].

The functional significance of spatial compartmentalization was further validated by genetic perturbation of nuclear architecture. In mammalian cell culture systems, transgenic tandem arrays of *E. coli lac* operator (*lacO*) sites can be tethered to the nuclear envelope by lacI proteins fused with Lamin B1 or lamin-associated Emerin and Lap2β proteins [37, 38]. This perinuclear tethering down-regulates the expression of reporter genes close to the engineered *lacO* clusters [37, 38]. In yeast, the partial disruption of a *cis*-acting silencer element near the mating-type *HMR* locus alleviates its transcriptional repression, but silencing is restored when HMR is artificially recruited to the nuclear envelope [39]. However, this restoration was not successful in a genetic background with disrupted foci of repressive SIR (silent information regulator) factors, indicating that being anchored to the nuclear envelope facilitates silencing, but that this epigenetic regulation requires additional factors or pathways [40].

Nuclear pore: in and beyond the nucleoplasm-cytoplasm transportation

The nuclear pore complexes (NPC) are molecular channels that pass through the nuclear envelope. They are permeable to small molecules but highly selective for macromolecules. Transport of large molecules depends on nuclear transport factors (NTFs) that bind to the transport signals, such as the short amino-acid sequences of nuclear localization signals (NLSs) and nuclear export signals (NESs) in cargo

proteins, to facilitate transport through the NPCs [41]. Various subtypes of NPCs exist and these are differentially distributed among different cell types and at specific developmental stages to achieve a diverse selectivity in nucleus-cytoplasmic transport [41].

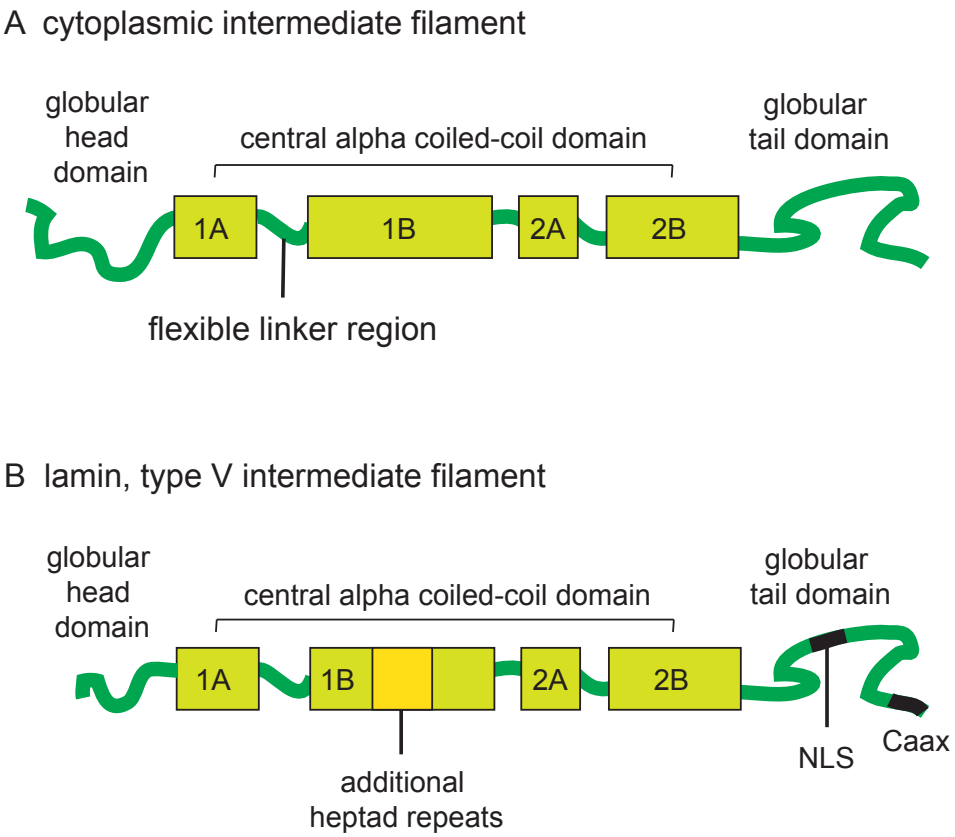
NPC-mediated transport also can exert an effect on nuclear morphology. The differences in nuclear size between the pseudotetraploid frog *Xenopus laevis* and its relative with a smaller diploid genome (*Xenopus tropicalis*) was mimicked in *Xenopus* egg extracts by altering the relative amounts of importin α and Ntf2, which are critical for lamin B3 import into the nucleus [42]. Also, Xenopus oocytes do not export a specific type of actin, resulting in its nuclear accumulation and the formation of actin filaments, which might be important for stabilization of the giant nucleus in this cell type [43].

Beyond the conventional function of nuclear-cytoplasmic transport, the NPCs seem to be the hubs of a multifunctional network on both sides of the nuclear envelope. On the cytoplasmic side, NPCs are connected to the cytoskeleton and translation machinery to coordinate messenger RNP (mRNP) export and the initiation of protein synthesis [44]. On the nucleoplasmic side, a network, which consists of macromolecular complexes and ribonucleoproteins, spreads from the baskets of NPCs into the nuclear periphery to connect to the neighboring NPCs. In animals, this network is referred to as the nuclear lamina, which plays important roles in maintaining the structural integrity of the nucleus as well as regulating gene expression and genomic integrity [45].

The nuclear lamina and the nucleoskeleton system in animals and non-plants

The nuclear lamina network in animal cells is largely composed of lamin proteins. These proteins were originally identified as the major components of the insoluble nuclear residue resistant to salt and detergent extraction [46]. Lamins are type V intermediate filament (IF) proteins with a conserved tri-partite structure, consisting of an N-terminal globular domain, a central coiled-coil rod domain, and a C-terminal globular domain (Figure 1.2, lamin A protein and IF protein) [47]. The rod domain is comprised of four coiled-coils (*i.e.*, 1A, 1B, 2A, 2B), and a conserved 42 amino-acid motif within 1B distinguishes type V from other IF proteins [48].

Long coiled-coil domains are frequently involved in oligomerization. The coils consist of multiple copies of a core heptad repeat with hydrophobic amino acids on the first (a) and fourth (d) position, often isoleucine, leucine or valine. These hydrophobic residues coil around the helix to form an amphipathic secondary structure [49]. In a hydrophilic cell environment, two such helices wrap around each other to bury the hydrophobic surfaces, which drives the oligomerization. The tight packing in a coiled-coil interface is stabilized by van der Waals contact between the side chains of a and d residues [49, 50]. This type of oligomerization has been shown for many cases of long coiled-coil domain proteins, such as SMC (Structural Maintenance of Chromosomes) proteins, tropomyosins, and lamin proteins [51-53]. *In vitro* experiments have shown that lamin proteins form dimers and then polymerize into higher-order filaments through lateral associations of dimers [51]. The lattice of lamin-based filaments



Adapted from CJ Hutchison *et al.* Nature Cell Biology 2004

1A, 1B, 2A, 2B: four major long alpha coiled-coils

NLS: Nuclear Localization Signal

Caax: a site for carboxyl methylation, farnesylation and proteolytic cleavage

Figure 1.2 Tripartite structures of intermediate filament protein and lamin protein

Figure 1.2 This cartoon illustrates the tripartite structure of intermediate filament protein and lamin protein. Figure 1.2A shows the general composition of an intermediate filament protein, consisting of a variable globular head domain, a long central alpha coiled-coil rod domain, and a globular tail domain. The central coiled-coil domain contains four major long coils, named 1A, 1B, 2A and 2B. Figure 1.2B shows the structure of lamin, a type V intermediate filament, where the N terminal globular head domain is short, 1B coil domain contains six heptad repeats, and globular tail domain includes a nuclear localization signal as well as CAAX site for farnesylation and proteolytic cleavage and carboxyl methylation during lamin protein maturation.

provides a cage for the mechanical support of the nuclei, as well as a platform for the interaction with a number of regulatory proteins [45].

In non-human metazoa, lamins are important determinants of nuclear organization. A variety of different experiments in experimental animal systems have been used to manipulate lamins to monitor the effect on chromatin and nuclear structure. For example, RNAi knockdown of the only gene encoding lamin in *C. elegans, Ce-lamin,* leads to pleiomorphic nuclei [54]. In mice, the ectopic expression of the germ-line-specific lamin, B3, in somatic cells causes the development of hook-shaped nuclei, similar to the nuclear structure seen in spermatocytes [55]. Expression of a dominant negative version of human lamin B1 in somatic cells causes dramatic deformation in the nuclear envelope [56]. A lamin-A mutation in cultured fibroblasts alters nuclear morphology and re-distributes histone modification marks for silenced heterochromatin, which appears to lose its association with the nuclear periphery [45]. Work in cell-free systems supports the conclusions of the genetic experiments discussed above. For instance, incubation of condensed chromatin together with lamin-depleted extracts from Xenopus oocytes results in small and fragile nuclei [57].

In human cells, lamin proteins also form a meshwork structure underneath the inner nuclear envelope. This nuclear lamina extends from nuclear pore to pore, as well as throughout the nucleoplasm, and binds a large number of specific genomic domains and lamin-associated proteins, such as emerin (encoded by the *EMD* gene, null mutations lead to Emery-Dreifuss muscular Dystrophy), LBR (Lamin B Receptor)

and LAP2 (Lamin Associated Polypeptide 2) [45]. Mutations and polymorphisms in lamin and lamin-associated proteins lead to malfunctioning nuclei and impair normal development. More than two hundred polymorphisms in the *LMNA* gene, which encodes lamin A and C, were discovered to be associated with a variety of human diseases collectively called laminopathies [58]. For instance, specific polymorphisms in *LMNA* cause the premature aging disease Hutchinson-Gilford Progeria Syndrome (HGPS). Cells isolated from laminopathy patients display fragile and irregularly-shaped nuclei with redistributed histone modification marks [58]. These observations suggest that a properly assembled nuclear lamina is important to maintain nuclear morphology and function, such as epigenetic modification and transcriptional regulation.

Plant, fungi and unicellular organisms do not have obvious lamin protein homologs. In yeast, the nucleus lack lamins. The functional compartmentalization in yeast nuclei is achieved by specific sequence elements, protein-protein interactions, nuclear envelope and nuclear pore anchorage sites, as well as chromosomal long-range interactions [59]. In the unicellular Trypanosomes, a large coiled-coil protein NUP-1 (nuclear peripheral protein 1) locates to the inner face of the nuclear envelope, as part of a stable network [60]. NUP-1 knockdown changes nuclear shape, disrupts the organization of nuclear pore complexes, and alters chromatin states, especially in the telomere-proximal region where silenced variant surface glycoprotein (VSG) genes are located. A loss of silencing in this region can increase VSG switching, which

allows the parasite to escape detection by the host immune systems. Therefore, NUP-1 is hypothesized to function like a lamin analog in this non-metazoan system [61].

Nuclear organization in Arabidopsis

Here, I review the current understanding of three aspects of nuclear organization in *Arabidopsis thaliana*, using representative studies: 1) nuclear proteins important for organelle structure; 2) the chromosome and chromatin scaffolding system; and 3) chromatin remodeling associated with dynamic nuclear organization. The proteins essential for plant nuclear organization are summarized in Table 1.1.

Table 1.1 A summary of nuclear organization relevant proteins in *A. thaliana*

proteins	location	function	mutants	nuclear morphology
SUN	Nuclear Envelope	SUN domain	*sun1sun2*	round
WIP	Nuclear Envelope	KASH domain	*wip1 wip2 wip3*	round
CRWN1/LINC1	Nuclear Envelope	nuclear and chromosome organization	*crwn1*	small and round not changed or slightly
CRWN2/LINC2	Nuclear Envelope	nuclear and chromosome organization	*crwn2*	rounder / smaller
CRWN3/LINC3	Nuclear Envelope	nuclear and chromosome organization	*crwn3*	not chnaged round with irregular
CRWN4/LINC4	Nuclear Envelope	nuclear and chromosome organization	*crwn4*	margin
GIP	Nuclear Envelope	nuclear-cytoplasmic connection	*gip1 gip2*	round
WIT	Nuclear Envelope	nuclear positioning and movement	*wit1 wit2*	invaginated NE and impaired nuclear movement
Xi-i	Nuclear Envelope	nuclear positioning and movement	*kaku1*	round
Nup136	Nuclear Pore Complex	nucleoporin / nuclear shape	*nup136 nup1*	round
Nup1	Nuclear Pore Complex	nucleoporin / nuclear shape	*nup136 nup1*	round
LNO1	Nuclear Pore Complex	nucleoporin / mRNA export	*lno1*	
MOS7	Nuclear Pore Complex	nucleoporin / immune response	*mos7*	
MOS3	Nuclear Pore Complex	nucleoporin / immune response	*mos3*	
Nup160	Nuclear Pore Complex	nucleoporin / mRNA export	*nup160*	
MFP1	S/MAR interacting protein	chromatin anchorage	*mfp1*	
MAF1	S/MAR interacting protein	chromatin anchorage	*maf1*	
AHL22	AT Hook S/MAR interacting protein	chromatin remodeling	*ahl22*	decondensed chromatin
TEK	AT Hook S/MAR interacting protein	chromatin remodeling	*tek*	decondensed chromatin
ARP4	Actin Related Protein in the nucleus	chromatin remodeling	*arp4*	
ARP5	Actin Related Protein in the nucleus	chromatin remodeling	*arp5*	
ARP6	Actin Related Protein in the nucleus	chromatin remodeling	*arp6*	
ARP7	Actin Related Protein in the nucleus	chromatin remodeling	*arp6*	
SMC1/3	Structural Maintainence of Chromosomes	cohesin, chromosome organization	*smc*	
SMC2/4	Structural Maintainence of Chromosomes	condensin, chromosome organization	*smc*	
SMC5/6	Structural Maintainence of Chromosomes	cohesion and HR mediated DNA repair	*smc*	
DMS3	SMC like, chromatin remodeling	RNA directed DNA methylation (RdDM)	*dms3*	decondensed chromatin
DDM1	chromatin remodeling	maitainence of DNA methylation	*ddm1*	decondensed chromatin
NRPD	chromatin remodeling	Pol IV sub unit, RdDM	*nrpd1/2*	decondensed chromatin
HDA6	chromatin remodeling	histone acetylation	*hda6*	decondensed chromatin

Table 1.1 This table summarizes nuclear proteins contributing to the nuclear organization in *Arabidopsis thaliana* mentioned in this chapter. The gene names, locus ID, morphological alterations in the absence of corresponding gene, are listed.

Arabidopsis nuclear proteins important for organelle structure

Plants appear to have evolved an independent nucleoskeletal system divergent from other eukaryotes [62]. Among the nucleoskeletal components in animals, very few of them have homologous proteins in plants (Table 1.1). One example is Sad1/UNC-84 (SUN)-domain proteins. There are five homologs in *Arabidopsis thaliana*: AtSUN1 and AtSUN2 contain the highly conserved C-terminal SUN domain, resembling SUN protein structure in yeast, worms, flies and mammals; while AtSUN3 – AtSUN5 harbor a SUN domain in the middle of the proteins [63]. Fluorescently tagged protein fusions for AtSUN1 and AtSUN2 form homo- and hetero-dimers, which localize to the inner nuclear envelope and show low mobility *in vivo* [64, 65]. The inner nuclear membrane AtSUN proteins interact with the outer nuclear membrane AtWIP(WPP domain-interacting proteins) proteins to form SUN-KASH (Klarsicht/ANC-1/Syne homology) complexes in Arabidopsis, as the linkers of the nucleoskeleton to the cytoskeleton (LINC). This LINC complex is conserved from yeast to human, and is involved in the positioning of nuclei and chromatin. The *sun1 sun2* double mutant has round nuclei, while the *wip1 wip2 wip3* triple mutants partially disrupt the spindle shape of the nuclei [65]. These findings indicate that the interaction between AtWIPs and AtSUNs are required to maintain the elongated nuclear shape typically seen in enlarged Arabidopsis cells [65].

Besides SUN-KASH proteins, a newly characterized GIP protein is also hypothesized to be involved in the nucleo-cytoplasmic connection [66]. GIP (γ-TuC protein 3

Interacting Protein 1) proteins are small components of γ-Tubulin Complexes (γ-TuCs) important for perinuclear localization and nucleation of microtubules (MTs). *gip1 gip2* mutants exhibited a severe alteration of nuclear shape, associated with an abnormal distribution of the NPCs and AtSUN1. Further, AtGIP1 also interacts with the nuclear envelope protein AtTSA1 [66].

Another type of nucleo-cytoplasmic linker consists of a plant-specific myosin motor XI-i and the nuclear membrane proteins WIT1 and WIT2, which might enable rapid nuclear positioning in response to environmental stimuli [67, 68]. XI-i is one of the thirteen myosin XI motor proteins responsible for transporting plant organelles on the actin cytoskeleton [69]. Myosin XI-i, which is encoded by the *KAKU1* gene in Arabidopsis, is anchored to the nuclear membrane by the outer nuclear membrane proteins WIT1 and WIT2. The *kaku1* mutant exhibits abnormal invagination of the nuclear envelope, and a deficiency of either myosin XI-i or WIT proteins diminishes dark-induced nuclear positioning in plant mesophyll cells [67].

Nuclear pore proteins implicated in nuclear organization

As in animal cells, the nuclear pore complexes (NPCs) primarily facilitate nucleo-cytoplasmic transport in plants. Proteomic studies showed that Arabidopsis contains at least 30 nucleoporins similar to nucleoporins in human and yeast. NUA (Nuclear Pore Anchor), an Arabidopsis ortholog of the long coiled-coil filament proteins of the nuclear pore basket, locates at the inner surface of the nuclear envelope and is

important in controlling mRNA export, as well as SUMO protease activity at the nuclear pore [70]. LNO1, homologous to Nup214 in human and Nup159 in yeast, encodes a nucleoporin protein required for mature mRNA export, and is essential for embryogenesis and seed viability [71]. MOS7, an Arabidopsis homolog to human and fly nucleoporin Nup88, localizes to the nuclear envelope and is required for the nuclear accumulation of SNC1, EDS1 and NPR1. *mos7-1* single mutant plants exhibit defects in basal and *R* protein-mediated immunity and in systemic acquired resistance [72]. In the same pathway, MOS3, an Arabidopsis homolog of Nup96 in vertebrates, is required for constitutive plant immunity mediated by the *R* gene *SNC1* [73]; while Arabidopsis Nup160 is needed for nuclear mRNA export and full expression of resistance dependent on EDS1 [74]. Additionally, a plant specific nucleoporin Nup136/Nup1 shows greater mobility on the nuclear envelope than other nucleoporins, and *nup136/nup1* mutants exhibit a rounder nuclear shape than wild type, as well as various defects in plant development [73].

CRWN family proteins as candidate molecular components of the Arabidopsis nucleoskeleton

In the 1990s, Masuda and colleagues reported the discovery of Nuclear Matrix Constituent Protein 1 (DcNMCP1) a protein residing on the periphery of carrot nuclei and a component of the salt-resistant nuclear matrix [75]. The localization of the protein and its structure, which featured a long coiled-coil domain, suggested that this protein might serve an architectural role in plant nuclei. This result was furthered

supported by the discovery of AcNMCP1 in onion (*Allium cepa*) and AgNMCP1 in celery (*Apium graveolens*), which are homologous to DcNMCP1 and belong to the nuclear matrix [76, 77]. Immuno-staining showed that AcNMCP1 proteins were concentrated around the nuclear rim, with some punctate signals in the nucleoplasm [75], while AgNMCP1 co-localized with the mitotic spindle and segregating chromosomes during mitosis [77]. NMCP1 and related proteins are plant-specific and share no significant amino-acid similarity to lamins, but their tripartite structure with a central coiled-coil domain is reminiscent of lamins. Moreover, the localization of these proteins at the nuclear periphery suggests that NMCP1-related plant proteins might function like lamin analogs [78, 79]. However, it is important to note that other structural analyses has suggested that NMCP proteins share more similarity with animal myosins or paramyosins than with lamins [80].

The Arabidopsis genome encodes four NMCP homologs, which were first identified in bioinformatic surveys of predicted proteins that contain coiled coil domains. Based on reverse genetics studies undertaken previously in the Richards lab and extended in this thesis, we refer to the Arabidopsis NMCP proteins as CRWN (CROWDED NUCLEI) based on the reduced nuclear size observed in *crwn* mutants [78] (also refer to Chapter 2]. These proteins are also known in the literature by their previous name, LINC (LITTLE NUCLEI) [78]. Our previous studies showed that the Arabidopsis CRWN1 and CRWN2 expressed as fluorescently-tagged fusion proteins in transgenic plants primarily localize to the nuclear periphery. The over-expression of CRWN2 leads to its distribution throughout the nucleoplasm, however [78]. A recent study

from another group shows that CRWN4 (LINC4) proteins are located at the nuclear periphery. CRWN1 (LINC1) proteins co-localize with mitotic chromosomes during the anaphase of cell cycle while the other CRWN proteins were distributed throughout the cell during mitosis [81].

Previous work in the Richards lab demonstrated that *CRWN1* and *CRWN2* play important roles in specifying nuclear shape and size, and *crwn1 crwn2* double mutants have reduced numbers of chromocenters, conspicuous heterochromatin aggregates visible at interphase [78]. Recently, another group also reported that disruption of *CRWN4* reduced nuclear size and caused loss of the elongated nuclear shape in differentiated cells [81]. Chapter 2 of this thesis presents a comprehensive study of the phenotypic effects caused by mutations in the *CRWN* gene family.

Scaffold proteins interacting with chromosomes

A genome scale analysis in Arabidopsis demonstrates that genes containing intragenic S/MARs predicted by SMARTest [82] are enriched for genes encoding transcription factors and are associated with more pronounced regulation (in contrast to constitutively expressed 'housekeeping' genes) [83]. These observations suggest a role for chromatin structural characteristics in mediating transcriptional regulation in a tissue-specific manner [83]. Transgenic constructs with strong MAR sequences flanking reporter genes stimulated median gene expression by five- to ten-fold in *Arabidopsis* [84]. Proteins associated with chromatin at MAR regions are considered

to be important for dynamic chromatin organization in various fundamental nuclear processes in all eukaryotic cells. An investigation of isolated plant nuclear matrix identified a candidate structural protein MFP1 (MAR binding filament-like protein 1) that is located at the nuclear periphery and is associated with speckle-like structures via its predicted N-terminal trans-membrane domain [85]. MFP1, a nuclear and plastid protein [86], binds to MAR DNA and may function in attaching the chromosomes to the nuclear envelope [85, 87]. Another small and soluble serine/threonine-rich protein MAF1 (MFP1-associated factor 1) also localizes to the nuclear periphery and interacts with MFP1 [88]. These two proteins are hypothesized to be candidate components of plant nuclear matrices [85].

Nuclear localized AT-hook proteins

Another emerging group of candidate nuclear structural proteins are AHL proteins (AT-Hook motif containing nuclear Localized), which contain two AT-hook motifs and one PPC (plant and prokaryote conserved) domain. These proteins localize to the nucleus and bind to AT-rich DNA sequences, such as S/MAR regions. There are 29 *AHL* genes in the *Arabidopsis* genome [89]. The AT-hook protein AHL22 (AT-hook motif nuclear localized 22) regulates flowering time by modifying chromatin at the *FT* gene (Flowering Locus T) [90]. Over expression of the *AHL22* gene reduces histone H3 acetylation and elevates histone H3K9 di-methylation, and activates the *FT* gene, resulting in delayed flowering [89, 90]. Another AT-hook DNA binding protein, TEK (Transposable Element silencing via AT-hooK), participates in silencing transposable

elements (TE) and TE-like sequence containing genes, such as Ler *FLC* (Flowering Locus C), *MAF4* (MADS Affecting Flowering 4), *MAF5*, *AtMu1* and *FWA*. Knockdown of TEK using miRNA leads to expression of these genes and reactivation of transposable elements, associated with increased histone acetylation, reduced histone H3K9 di-methylation, and DNA hypomethylation in these loci [91-93].

Nuclear localized Actin Related proteins (ARP)

The actin-related proteins (ARPs) that locate to the nucleus are also involved in dynamic chromatin organization and transcriptional regulation. In *Arabidopsis*, there are four ARP proteins, and all of them are present in chromatin modifying complexes. AtARP4, a conserved homolog of human BAF53 and yeast Arp4, belongs to chromatin modifying complexes and is concentrated in the nucleoplasm of plant cells. The *arp4-1* mutant, which exhibits reduced ARP protein activity, has defects in anther development and is partially sterile. A complete knockdown of ARP4 by RNAi led to strong pleiotropic phenotypes in organ organization, timing of flowering, and high levels of sterility [94]. Another ARP member in *Arabidopsis*, AtARP5, corresponds to a conserved subunit of the INO80 chromatin-remodeling complex in yeast and mammals. The AtARP5 protein is ubiquitously expressed and localizes to the nucleoplasm of interphase cells. A null *arp5* mutant produces moderately dwarfed plants, with reduced cell size and delayed stomatal development. These plants are also hypersensitive to DNA-damaging reagents [95]. Moreover, AtARP6, a homolog of ARP6 in yeast and other organisms, is a component of the SWR1 chromatin-

remodeling complex. *AtARP6*is also universally expressed and localizes to the nucleus in interphase. Null mutations in *AtARP6* caused numerous developmental defects, female fertility and early flowering. Some of these phenotypic changes could be explained by a down-regulation of *FLC* (Flowering Locus C), *MAF4* and *MAF5* (MADS Affecting Flowering 4 and 5) in *arp6* mutants [96, 97]. AtARP7 is a constitutively expressed nuclear protein. The *arp7-1* T-DNA mutant is homozygous lethal and produces abnormal homozygous embryos that are blocked in their development at or before the torpedo stage. A drastic knockdown of *AtARP7* by RNAi results in severe dwarfism, pleiotropic phenotypes, and defective cell expansion and trichome morphology [98]. It has been proposed that these transcriptional and developmental changes in various *arp* mutants are a result of transcriptional mis-regulation mediated by altered chromatin structures due to the loss of ARP proteins in the chromatin modifying complexes [99].

Structural Maintenance of Chromosome (SMC) complexes

Similar to other eukaryotes, plants use SMC (Structural Maintenance of Chromosome) protein complexes and their interacting partners for sister chromatid cohesion and chromosome condensation [100]. Arabidopsis contains single copies of the cohesion subunits SMC1 and SMC3. Both of these subunits are required for sister chromatid cohesion along chromosome arms in somatic interphase nuclei [100]. Condensin subunits include SMC2A/B and SMC4A/B, which are important for chromosome

condensation in mitosis and meiosis [100]. Besides scaffolding the chromosomes, the SMC complexes also participate in other basic nuclear functions, such as DNA repair and epigenetic regulation. In *Arabidopsis*, the incompletely aligned sister-chromatids in interphase nuclei increase the chromatin accessibility and facilitate frequent inter-chromosomal recombination [101]. In somatic cells, this feature is important for efficient DNA repair via homologous recombination guided by the SMC5/6 complex [102]. Moreover, *DMS3* (*At3g49250*) encodes a protein similar to the hinge-domain region of SMC proteins, which functions together with RDR1 and Pol IVb subunits in the RdDM (RNA-directed DNA methylation) machinery. These complexes establish cytosine methylation on DNA regions homologous to the Dicer-generated small RNAs to achieve gene silencing [103]. These examples suggest that chromosome architectural proteins play various functional roles in maintaining the chromosome structure and function.

Chromatin remodeling and dynamic nuclear organization in Arabidopsis

Epigenetic reprogramming during plant development is also accompanied by significant alteration in chromatin organization. At the beginning of seed maturation, embryonic cotyledon nuclei display significantly reduced nuclear size and highly condensed chromatin [104]. These morphologies are released during germination, when an increase in nuclear size occurs, along with a partial decondensation of the chromatin and restoration of transcription [104]. In the floral transition, chromosomes go through another round of decondensation and transcriptional reprogramming to

accomplish developmental changes [105]. Moreover, in experimentally cultured plant cells, dedifferentiation of mesophyll cells into protoplasts disrupts chromocenters and decondenses many major heterochromatic repeat regions. Under specific conditions, this process could be reversed by re-condensation of heterochromatin into chromocenters in a stepwise manner [106].

Several layers of epigenetic remodeling are associated with reorganization of chromatin compaction in the nuclei. Examples include DNA methylation, small RNA silencing, histone modification, and crosstalk among these pathways [106]. DNA methylation plays important roles in transcriptional silencing and heterochromatin maintenance. The *ddm1* mutant in *Arabidopsis,* with its defect in the maintenance of DNA methylation, shows de-condensed chromocenters in the nuclei of mature cells [106]. Small interfering RNAs also participate in this silencing process via siRNA-directed DNA methylation. The *Arabidopsis* null mutants for plant-specific RNA polymerases IV impair the biogenesis of siRNAs, release the silencing of transposable elements and pericentromeric repeats, and cause the dispersal of heterochromatin [107]. Histone acetylation and deacetylation are also involved in mediating transcriptional activation and silencing. *Arabidopsis* plants carrying a mutation in *HDA6*, a class I histone deacetylase gene, is associated with increased histone H3K4 methylation, decondensation of the chromatin, and a loss of transcriptional silencing at several transgenic loci as well as in endogenous repetitive regions [108].

The re-organization of heterochromatin in *Arabidopsis* can also be induced by

environmental cues. One example comes from the exploration of *Arabidopsis* natural variation, which revealed a positive correlation among chromatin compaction, latitude of geographic origin and local light intensity. The natural strain Cape Verde Islands-0 (Cvi-0) contains a polymorphism in the *PHYB* gene and within the *HDA6* promoter. This genotype leads to a plant that resembles either *hda6* or *phyB5* mutant in having reduced chromatin compaction [110]. Cvi-0 also displayed decreased methylation levels of DNA and histone H3 K9 methylation at the ribosomal RNA gene clusters as the loss of HDA6 protein does [110]. Thus both PHYB (Phytochrome-B) and HDA6 promote global chromatin compaction level in response to light, suggesting that acclimation of *Arabidopsis* to its environment could be associated with chromatin plasticity [109]. In another example, prolonged heat stress showed transcriptional activation of several repetitive elements in the heterochromatic region that are under epigenetic regulation. These gene expression changes were accompanied by loss of nucleosomes and by heterochromatin decondensation. However, this change was transient, and did not involve clear repressive epigenetic changes such as DNA hypomethylation and changes in histone modifications. The nucleosome loading and transcriptional silencing was reestablished after the plant recovered from heat stress, but a delayed recovery was observed in mutants with chromatin assembly deficiencies [110]. Therefore, chromatin organization can react to environmental cues independent of epigenetic modifications, and this feature might facilitate more permanent epigenetic changes [110].

Summary

Architecture is something that makes space meaningful. The same thing holds true for nuclei. Certain structures were needed to define the margin of the first nucleus. Once formed, the nucleus and its organization started to evolve and pick up new functions along the evolutionary path. Progress in the study of nuclear organization - the dynamic architectural components and various molecular machinery functioning in that environment - promises to uncover a new layer of spatial regulation of nuclear functions in both non-plant and plant systems.

Overview of Dissertation

In this dissertation, I took advantage of *Arabidopsis thaliana* as a model system in plants, to study nuclear organization and its impact on nuclear function. The work focuses on a plant-specific long coiled-coil domain protein family, CRWN, which stands for "CRoWded Nuclei".

Chapter 2 contains a description of my genetic analysis of all the viable *crwn* mutant combinations. Phenotypes I observed in these mutants included whole plant dwarfism, nuclear size reduction, increased nuclear DNA density, and disruption of heterochromatin organization. The phylogenetic analysis of CRWN proteins, as well

as the morphological alterations I recorded on both the whole plant and nuclear levels point to a functional divergence between CRWN1-like and CRWN4 proteins in the family. Chapter 2 has been published in *BMC Plant Biology* 2013, 13:200.

In Chapter 3, I performed mRNA-seq profiling on six select *crwn* mutant genotypes. This analysis confirmed the synergistic relationship among CRWN1, CRWN2 and CRWN3 proteins, and uncovered a suppression between *crwn1* and *crwn4* mutations in *crwn1 crwn4* double mutants. Moreover, the most severe mutants, *crwn4* and *crwn1 crwn2*, shared common genomic targets in transcriptional mis-regulation. Many mis-expressed genes were identified as candidates for corresponding morphological changes in *crwn* mutants. Thes mutants also display a mild release of epigenetic silencing and activation of various stress response pathways. Many structural and functional nuclear proteins were mis-expressed, potentially leading to disruption of both nuclear organization and basic nuclear processes, such as DNA replication, epigenetic modification, and DNA repair.

In Chapter 4, I obtained antisera specifically recognizing CRWN1 or CRWN4 proteins. Cell fractionation experiments confirmed that, as like other NMCP proteins, both CRWN1 and CRWN4 proteins are resistant to high salt and mild detergent treatment, thus could be candidate components of nucleoskeleton. Interestingly, I demonstrated that CRWN4 protein is down-regulated in *crwn1* mutants. Also, a compensatory regulation among different *CRWN* genes exists on the mRNA level,

suggesting a complex set of interactions among these paralogs. Last, I used co-immunoprecipitation to demonstrate that CRWN1 and CRWN4 proteins physically interact *in vivo*, supporting a model that CRWN1-like and CRWN4 proteins are working together to organize plant nuclei. I hypothesize that a balance between the divergent CRWN1-like and CRWN4 function is essential for proper nuclear organization.

Chapter 5 contains a brief summary of this thesis and a discussion of potential future directions.

REFERENCES

1. Dobell C, Leeuwenhoek Av: **Antony van Leeuwenhoek and his "Little animals"; being some account of the father of protozoology and bacteriology and his multifarious discoveries in these disciplines**. New York,: Harcourt, Brace and company; 1932.
2. Harris H: **The birth of the cell**. New Haven, Conn.: Yale University Press; 1999.
3. Brown R: **On the organs and mode of fecundation and Orchidex and Asclepiadea**. In: *Miscellaneous Botanical Works I: 511-514.* 1866.
4. Cremer T: **Von der Zellenlehre zur Chromosomentheorie : naturwissenschaftliche Erkenntnis und Theorienwechsel in der frühen Zell- und Vererbungsforschung**. Berlin ; New York: Springer-Verlag; 1985.
5. Lukacs D: **[Walter Flemming, discoverer of chromatin and mitotic cell division]**. *Orvosi hetilap* 1981, **122**(6):349-350.
6. Rabl C: **"Uber Zelltheilung" Morphologisches Jahrbuch**. 1885, **10**.
7. E H: **Das Heterochromatinder Moose. I**. *Jahrh wiss Bot* 1928, **69**:762-818.
8. Solovei I, Kreysing M, Lanctot C, Kosem S, Peichl L, Cremer T, Guck J, Joffe B: **Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution**. *Cell* 2009, **137**(2):356-368.
9. Pyrpasopoulou A, Meier J, Maison C, Simos G, Georgatos SD: **The lamin B receptor (LBR) provides essential chromatin docking sites at the nuclear envelope**. *The EMBO journal* 1996, **15**(24):7108-7119.

10. Ye Q, Callebaut I, Pezhman A, Courvalin JC, Worman HJ: **Domain-specific interactions of human HP1-type chromodomain proteins and inner nuclear membrane protein LBR**. *The Journal of biological chemistry* 1997, **272**(23):14983-14989.

11. Ye Q, Worman HJ: **Interaction between an integral protein of the nuclear envelope inner membrane and human chromodomain proteins homologous to Drosophila HP1**. *The Journal of biological chemistry* 1996, **271**(25):14653-14656.

12. Makatsori D, Kourmouli N, Polioudaki H, Shultz LD, McLean K, Theodoropoulos PA, Singh PB, Georgatos SD: **The inner nuclear membrane protein lamin B receptor forms distinct microdomains and links epigenetically marked chromatin to the nuclear envelope**. *The Journal of biological chemistry* 2004, **279**(24):25567-25573.

13. Hirano Y, Hizume K, Kimura H, Takeyasu K, Haraguchi T, Hiraoka Y: **Lamin B receptor recognizes specific modifications of histone H4 in heterochromatin formation**. *The Journal of biological chemistry* 2012, **287**(51):42654-42663.

14. Muller HJ: **Types of visible variations induced by X-rays in Drosophila**. *Journal of Genetics* 1930, **22**(3):299-334.

15. Wallrath LL, Elgin SC: **Position effect variegation in Drosophila is associated with an altered chromatin structure**. *Genes & development* 1995, **9**(10):1263-1277.

16. Dorer DR, Henikoff S: **Expansions of transgene repeats cause heterochromatin formation and gene silencing in Drosophila**. *Cell* 1994, **77**(7):993-1002.

17. Duboule D: **The rise and fall of Hox gene clusters**. *Development* 2007, **134**(14):2549-2560.

18. Caron H, van Schaik B, van der Mee M, Baas F, Riggins G, van Sluis P, Hermus MC, van Asperen R, Boon K, Voute PA *et al*: **The human transcriptome map: clustering of highly expressed genes in chromosomal domains**. *Science* 2001, **291**(5507):1289-1292.

19. Mirkovitch J, Mirault ME, Laemmli UK: **Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold**. *Cell* 1984, **39**(1):223-232.

20. Heng HH, Goetze S, Ye CJ, Liu G, Stevens JB, Bremer SW, Wykes SM, Bode J, Krawetz SA: **Chromatin loops are selectively anchored using scaffold/matrix-attachment regions**. *Journal of cell science* 2004, **117**(Pt 7):999-1008.

21. Cai S, Lee CC, Kohwi-Shigematsu T: **SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes**. *Nature genetics* 2006, **38**(11):1278-1288.

22. Wang L, Di LJ, Lv X, Zheng W, Xue Z, Guo ZC, Liu DP, Liang CC: **Inter-MAR association contributes to transcriptionally active looping events in human beta-globin gene cluster**. *PLoS One* 2009, **4**(2):e4629.

23. Dietzel S, Jauch A, Kienle D, Qu G, Holtgreve-Grez H, Eils R, Munkel C, Bittner M, Meltzer PS, Trent JM *et al*: **Separate and variably shaped chromosome arm domains are disclosed by chromosome arm painting in human cell nuclei**. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* 1998, **6**(1):25-33.

24. Dekker J, Marti-Renom MA, Mirny LA: **Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data**. *Nature reviews Genetics* 2013, **14**(6):390-403.

25. Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO *et al*: **Comprehensive mapping of long-range interactions reveals folding principles of the human genome**. *Science* 2009, **326**(5950):289-293.

26. Schoenfelder S, Sexton T, Chakalova L, Cope NF, Horton A, Andrews S, Kurukuti S, Mitchell JA, Umlauf D, Dimitrova DS *et al*: **Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells**. *Nature genetics* 2010, **42**(1):53-61.

27. Lee TH, Kim SJ, Han YM, Yu DY, Lee CS, Choi YJ, Moon HB, Baik MG, Lee KK: **Matrix attachment region sequences enhanced the expression frequency of a whey acidic protein/human lactoferrin fusion gene in the mammary gland of transgenic mice**. *Molecules and cells* 1998, **8**(5):530-536.

28. Keaton MA, Taylor CM, Layer RM, Dutta A: **Nuclear scaffold attachment sites within ENCODE regions associate with actively transcribed genes**. *PloS one* 2011, **6**(3):e17912.

29. Hancock R: **A role for macromolecular crowding effects in the assembly and function of compartments in the nucleus**. *Journal of structural biology* 2004, **146**(3):281-290.

30. Walter A, Chapuis C, Huet S, Ellenberg J: **Crowded chromatin is not sufficient for heterochromatin formation and not required for its maintenance**. *Journal of structural biology* 2013, **184**(3):445-453.

31. Palladino F, Laroche T, Gilson E, Axelrod A, Pillus L, Gasser SM: **SIR3 and SIR4 proteins are required for the positioning and integrity of yeast telomeres**. *Cell* 1993, **75**(3):543-555.

32. Brickner JH, Walter P: **Gene recruitment of the activated INO1 locus to the nuclear membrane**. *PLoS biology* 2004, **2**(11):e342.

33. Bupp JM, Martin AE, Stensrud ES, Jaspersen SL: **Telomere anchoring at the nuclear periphery requires the budding yeast Sad1-UNC-84 domain protein Mps3**. *The Journal of cell biology* 2007, **179**(5):845-854.

34. Grund SE, Fischer T, Cabal GG, Antunez O, Perez-Ortin JE, Hurt E: **The inner nuclear membrane protein Src1 associates with subtelomeric genes and alters their regulated gene expression**. *The Journal of cell biology* 2008, **182**(5):897-910.

35. van Steensel B, Delrow J, Henikoff S: **Chromatin profiling using targeted DNA adenine methyltransferase**. *Nature genetics* 2001, **27**(3):304-308.

36. Orian A: **Chromatin profiling, DamID and the emerging landscape of gene expression**. *Curr Opin Genet Dev* 2006, **16**(2):157-164.

37. Reddy KL, Zullo JM, Bertolino E, Singh H: **Transcriptional repression mediated by repositioning of genes to the nuclear lamina**. *Nature* 2008, **452**(7184):243-247.

38. Kumaran RI, Spector DL: **A genetic locus targeted to the nuclear periphery in living cells maintains its transcriptional competence**. *The Journal of cell biology* 2008, **180**(1):51-65.

39. Taddei A, Gasser SM: **Multiple pathways for telomere tethering: functional implications of subnuclear position for heterochromatin formation**. *Biochimica et biophysica acta* 2004, **1677**(1-3):120-128.

40. Taddei A, Van Houwe G, Nagai S, Erb I, van Nimwegen E, Gasser SM: **The functional importance of telomere clustering: global changes in gene expression result from SIR factor dispersion**. *Genome research* 2009, **19**(4):611-625.

41. Strambio-De-Castillia C, Niepel M, Rout MP: **The nuclear pore complex: bridging nuclear transport and gene regulation**. *Nature reviews Molecular cell biology* 2010, **11**(7):490-501.

42. Levy DL, Heald R: **Nuclear size is regulated by importin alpha and Ntf2 in Xenopus**. *Cell* 2010, **143**(2):288-298.

43. Feric M, Brangwynne CP: **A nuclear F-actin scaffold stabilizes ribonucleoprotein droplets against gravity in large cells**. *Nature cell biology* 2013, **15**(10):1253-1259.

44. Akhtar A, Gasser SM: **The nuclear envelope and transcriptional control**. *Nature reviews Genetics* 2007, **8**(7):507-517.

45. Dechat T, Pfleghaar K, Sengupta K, Shimi T, Shumaker DK, Solimando L, Goldman RD: **Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin**. *Genes & development* 2008, **22**(7):832-853.

46. Aaronson RP, Blobel G: **Isolation of nuclear pore complexes in association with a lamina**. *Proceedings of the National Academy of Sciences of the United States of America* 1975, **72**(3):1007-1011.

47. Hutchison CJ, Worman HJ: **A-type lamins: guardians of the soma?** *Nature cell biology* 2004, **6**(11):1062-1067.

48. Dechat T, Adam SA, Taimen P, Shimi T, Goldman RD: **Nuclear lamins**. *Cold Spring Harbor perspectives in biology* 2010, **2**(11):a000547.

49. Mason JM, Arndt KM: **Coiled coil domains: stability, specificity, and biological implications**. *Chembiochem : a European journal of chemical biology* 2004, **5**(2):170-176.

50. Hanukoglu I, Ezra L: **Proteopedia: Coiled-coil structure of keratins**. *Biochemistry and molecular biology education : a bimonthly publication of the International Union of Biochemistry and Molecular Biology* 2013.

51. Strelkov SV, Schumacher J, Burkhard P, Aebi U, Herrmann H: **Crystal structure of the human lamin A coil 2B dimer: implications for the head-to-tail association of nuclear lamins**. *Journal of molecular biology* 2004, **343**(4):1067-1080.

52. Hirano T: **At the heart of the chromosome: SMC proteins in action**. *Nature reviews Molecular cell biology* 2006, **7**(5):311-322.

53. McLachlan AD, Stewart M: **Tropomyosin coiled-coil interactions: evidence for an unstaggered structure**. *Journal of molecular biology* 1975, **98**(2):293-304.

54. Gruenbaum Y, Lee KK, Liu J, Cohen M, Wilson KL: **The expression, lamin-dependent localization and RNAi depletion phenotype for emerin in C. elegans**. *Journal of cell science* 2002, **115**(Pt 5):923-929.

55. Furukawa K, Hotta Y: **cDNA cloning of a germ cell specific lamin B3 from mouse spermatocytes and analysis of its function by ectopic expression in somatic cells**. *The EMBO journal* 1993, **12**(1):97-106.

56. Butin-Israeli V, Ben-nun-Shaul O, Kopatz I, Adam SA, Shimi T, Goldman RD, Oppenheim A: **Simian virus 40 induces lamin A/C fluctuations and nuclear envelope deformation during cell entry**. *Nucleus* 2011, **2**(4):320-330.

57. Jenkins H, Holman T, Lyon C, Lane B, Stick R, Hutchison C: **Nuclei that lack a lamina accumulate karyophilic proteins and assemble a nuclear matrix**. *Journal of cell science* 1993, **106 (Pt 1)**:275-285.

58. Butin-Israeli V, Adam SA, Goldman AE, Goldman RD: **Nuclear lamin functions and disease**. *Trends in genetics : TIG* 2012, **28**(9):464-471.

59. Taddei A, Gasser SM: **Structure and function in the budding yeast nucleus**. *Genetics* 2012, **192**(1):107-129.

60. Rout MP, Field MC: **Isolation and characterization of subnuclear compartments from Trypanosoma brucei. Identification of a major repetitive nuclear lamina component**. *The Journal of biological chemistry* 2001, **276**(41):38261-38271.

61. DuBois KN, Alsford S, Holden JM, Buisson J, Swiderski M, Bart JM, Ratushny AV, Wan Y, Bastin P, Barry JD *et al*: **NUP-1 Is a large coiled-coil nucleoskeletal protein in trypanosomes with lamin-like functions**. *PLoS biology* 2012, **10**(3):e1001287.

62. Rose A, Patel S, Meier I: **Plant nuclear envelope proteins**. *Symposia of the Society for Experimental Biology* 2004(56):69-88.

63. Graumann K, Runions J, Evans DE: **Characterization of SUN-domain proteins at the higher plant nuclear envelope**. *The Plant journal : for cell and molecular biology* 2010, **61**(1):134-144.

64. Oda Y, Fukuda H: **Dynamics of Arabidopsis SUN proteins during mitosis and their involvement in nuclear shaping**. *The Plant journal : for cell and molecular biology* 2011, **66**(4):629-641.

65. Zhou X, Graumann K, Evans DE, Meier I: **Novel plant SUN-KASH bridges are involved in RanGAP anchoring and nuclear shape determination**. *J Cell Biol* 2012, **196**(2):203-211.

66. Janski N, Masoud K, Batzenschlager M, Herzog E, Evrard JL, Houlne G, Bourge M, Chaboute ME, Schmit AC: **The GCP3-interacting proteins GIP1 and GIP2 are required for gamma-tubulin complex protein localization,**

spindle integrity, and chromosomal stability. *The Plant cell* 2012, **24**(3):1171-1187.

67. Tamura K, Iwabuchi K, Fukao Y, Kondo M, Okamoto K, Ueda H, Nishimura M, Hara-Nishimura I: **Myosin XI-i links the nuclear membrane to the cytoskeleton to control nuclear movement and shape in Arabidopsis**. *Current biology : CB* 2013, **23**(18):1776-1781.

68. Brkljacic J, Zhao Q, Meier I: **WPP-domain proteins mimic the activity of the HSC70-1 chaperone in preventing mistargeting of RanGAP1-anchoring protein WIT1**. *Plant Physiol* 2009, **151**(1):142-154.

69. Sattarzadeh A, Schmelzer E, Hanson MR: **Arabidopsis myosin XI sub-domains homologous to the yeast myo2p organelle inheritance sub-domain target subcellular structures in plant cells**. *Frontiers in plant science* 2013, **4**:407.

70. Xu XM, Rose A, Muthuswamy S, Jeong SY, Venkatakrishnan S, Zhao Q, Meier I: **NUCLEAR PORE ANCHOR, the Arabidopsis homolog of Tpr/Mlp1/Mlp2/megator, is involved in mRNA export and SUMO homeostasis and affects diverse aspects of plant development**. *The Plant cell* 2007, **19**(5):1537-1548.

71. Braud C, Zheng W, Xiao W: **LONO1 encoding a nucleoporin is required for embryogenesis and seed viability in Arabidopsis**. *Plant physiology* 2012, **160**(2):823-836.

72. Cheng YT, Germain H, Wiermer M, Bi D, Xu F, Garcia AV, Wirthmueller L, Despres C, Parker JE, Zhang Y *et al*: **Nuclear pore complex component**

MOS7/Nup88 is required for innate immunity and nuclear accumulation of defense regulators in Arabidopsis. *The Plant cell* 2009, **21**(8):2503-2516.

73. Zhang Y, Li X: **A putative nucleoporin 96 Is required for both basal defense and constitutive resistance responses mediated by suppressor of npr1-1,constitutive 1**. *The Plant cell* 2005, **17**(4):1306-1316.

74. Wiermer M, Cheng YT, Imkampe J, Li M, Wang D, Lipka V, Li X: **Putative members of the Arabidopsis Nup107-160 nuclear pore sub-complex contribute to pathogen defense**. *The Plant journal : for cell and molecular biology* 2012, **70**(5):796-808.

75. Masuda K, Xu ZJ, Takahashi S, Ito A, Ono M, Nomura K, Inoue M: **Peripheral framework of carrot cell nucleus contains a novel protein predicted to exhibit a long alpha-helical domain**. *Experimental cell research* 1997, **232**(1):173-181.

76. Kimura Y, Kuroda C, Masuda K: **Differential nuclear envelope assembly at the end of mitosis in suspension-cultured Apium graveolens cells**. *Chromosoma* 2010, **119**(2):195-204.

77. Ciska M, Masuda K, Moreno Diaz de la Espina S: **Lamin-like analogues in plants: the characterization of NMCP1 in Allium cepa**. *Journal of experimental botany* 2013, **64**(6):1553-1564.

78. Dittmer TA, Stacey NJ, Sugimoto-Shirasu K, Richards EJ: **LITTLE NUCLEI genes affecting nuclear morphology in Arabidopsis thaliana**. *The Plant cell* 2007, **19**(9):2793-2803.

79. Ciska M, Moreno S: **NMCP/LINC proteins: Putative lamin analogs in plants?** *Plant signaling & behavior* 2013, **8**(12).

80. Gardiner J, Overall R, Marc J: **Putative Arabidopsis homologues of metazoan coiled-coil cytoskeletal proteins**. *Cell biology international* 2011, **35**(8):767-774.

81. Sakamoto Y, Takagi S: **LITTLE NUCLEI 1 and 4 regulate nuclear morphology in Arabidopsis thaliana**. *Plant & cell physiology* 2013, **54**(4):622-633.

82. Rudd S, Frisch M, Grote K, Meyers BC, Mayer K, Werner T: **Genome-wide in silico mapping of scaffold/matrix attachment regions in Arabidopsis suggests correlation of intragenic scaffold/matrix attachment regions with gene expression**. *Plant physiology* 2004, **135**(2):715-722.

83. Tetko IV, Haberer G, Rudd S, Meyers B, Mewes HW, Mayer KF: **Spatiotemporal expression control correlates with intragenic scaffold matrix attachment regions (S/MARs) in Arabidopsis thaliana**. *PLoS computational biology* 2006, **2**(3):e21.

84. Van der Geest AH, Welter ME, Woosley AT, Pareddy DR, Pavelko SE, Skokut M, Ainley WM: **A short synthetic MAR positively affects transgene expression in rice and Arabidopsis**. *Plant biotechnology journal* 2004, **2**(1):13-26.

85. Meier I, Phelan T, Gruissem W, Spiker S, Schneider D: **MFP1, a novel plant filament-like protein with affinity for matrix attachment region DNA**. *The Plant cell* 1996, **8**(11):2105-2115.

86. Jeong SY, Rose A, Meier I: **MFP1 is a thylakoid-associated, nucleoid-binding protein with a coiled-coil structure**. *Nucleic acids research* 2003, **31**(17):5175-5185.

87. Gindullis F, Meier I: **Matrix attachment region binding protein MFP1 is localized in discrete domains at the nuclear envelope**. *The Plant cell* 1999, **11**(6):1117-1128.

88. Gindullis F, Peffer NJ, Meier I: **MAF1, a novel plant protein interacting with matrix attachment region binding protein MFP1, is located at the nuclear envelope**. *The Plant cell* 1999, **11**(9):1755-1768.

89. Xiao C, Chen F, Yu X, Lin C, Fu YF: **Over-expression of an AT-hook gene, AHL22, delays flowering and inhibits the elongation of the hypocotyl in Arabidopsis thaliana**. *Plant molecular biology* 2009, **71**(1-2):39-50.

90. Yun J, Kim YS, Jung JH, Seo PJ, Park CM: **The AT-hook motif-containing protein AHL22 regulates flowering initiation by modifying FLOWERING LOCUS T chromatin in Arabidopsis**. *The Journal of biological chemistry* 2012, **287**(19):15307-15316.

91. Xu Y, Gan ES, He Y, Ito T: **Flowering and genome integrity control by a nuclear matrix protein in Arabidopsis**. *Nucleus* 2013, **4**(4):274-276.

92. Xu Y, Gan ES, Ito T: **The AT-hook/PPC domain protein TEK negatively regulates floral repressors including MAF4 and MAF5**. *Plant signaling & behavior* 2013, **8**(8).

93. Xu Y, Wang Y, Stroud H, Gu X, Sun B, Gan ES, Ng KH, Jacobsen SE, He Y, Ito T: **A matrix protein silences transposons and repeats through**

interaction with retinoblastoma-associated proteins. *Current biology : CB* 2013, **23**(4):345-350.

94. Kandasamy MK, Deal RB, McKinney EC, Meagher RB: **Silencing the nuclear actin-related protein AtARP4 in Arabidopsis has multiple effects on plant development, including early flowering and delayed floral senescence**. *The Plant journal : for cell and molecular biology* 2005, **41**(6):845-858.

95. Kandasamy MK, McKinney EC, Deal RB, Smith AP, Meagher RB: **Arabidopsis actin-related protein ARP5 in multicellular development and DNA repair**. *Developmental biology* 2009, **335**(1):22-32.

96. Choi K, Kim S, Kim SY, Kim M, Hyun Y, Lee H, Choe S, Kim SG, Michaels S, Lee I: **SUPPRESSOR OF FRIGIDA3 encodes a nuclear ACTIN-RELATED PROTEIN6 required for floral repression in Arabidopsis**. *The Plant cell* 2005, **17**(10):2647-2660.

97. Martin-Trillo M, Lazaro A, Poethig RS, Gomez-Mena C, Pineiro MA, Martinez-Zapater JM, Jarillo JA: **EARLY IN SHORT DAYS 1 (ESD1) encodes ACTIN-RELATED PROTEIN 6 (AtARP6), a putative component of chromatin remodelling complexes that positively regulates FLC accumulation in Arabidopsis**. *Development* 2006, **133**(7):1241-1252.

98. Kandasamy MK, McKinney EC, Deal RB, Meagher RB: **Arabidopsis ARP7 is an essential actin-related protein required for normal embryogenesis, plant architecture, and floral organ abscission**. *Plant physiology* 2005, **138**(4):2019-2032.

99. Meagher RB, Kandasamy MK, Smith AP, McKinney EC: **Nuclear actin-related proteins at the core of epigenetic control**. *Plant signaling & behavior* 2010, **5**(5):518-522.

100. Schubert V: **SMC proteins and their multiple functions in higher plants**. *Cytogenetic and genome research* 2009, **124**(3-4):202-214.

101. Schubert V, Klatte M, Pecinka A, Meister A, Jasencakova Z, Schubert I: **Sister chromatids are often incompletely aligned in meristematic and endopolyploid interphase nuclei of Arabidopsis thaliana**. *Genetics* 2006, **172**(1):467-475.

102. Watanabe K, Pacher M, Dukowic S, Schubert V, Puchta H, Schubert I: **The STRUCTURAL MAINTENANCE OF CHROMOSOMES 5/6 complex promotes sister chromatid alignment and homologous recombination after DNA damage in Arabidopsis thaliana**. *The Plant cell* 2009, **21**(9):2688-2699.

103. Kanno T, Bucher E, Daxinger L, Huettel B, Bohmdorfer G, Gregor W, Kreil DP, Matzke M, Matzke AJ: **A structural-maintenance-of-chromosomes hinge domain-containing protein is required for RNA-directed DNA methylation**. *Nature genetics* 2008, **40**(5):670-675.

104. van Zanten M, Koini MA, Geyer R, Liu Y, Brambilla V, Bartels D, Koornneef M, Fransz P, Soppe WJ: **Seed maturation in Arabidopsis thaliana is characterized by nuclear size reduction and increased chromatin condensation**. *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**(50):20219-20224.

105. Tessadori F, Schulkes RK, van Driel R, Fransz P: **Light-regulated large-scale reorganization of chromatin during the floral transition in Arabidopsis**. *The Plant journal : for cell and molecular biology* 2007, **50**(5):848-857.

106. Fransz P, De Jong JH, Lysak M, Castiglione MR, Schubert I: **Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(22):14584-14589.

107. Pikaard CS, Haag JR, Ream T, Wierzbicki AT: **Roles of RNA polymerase IV in gene silencing**. *Trends in plant science* 2008, **13**(7):390-397.

108. Probst AV, Fagard M, Proux F, Mourrain P, Boutet S, Earley K, Lawrence RJ, Pikaard CS, Murfett J, Furner I *et al*: **Arabidopsis histone deacetylase HDA6 is required for maintenance of transcriptional gene silencing and determines nuclear organization of rDNA repeats**. *The Plant cell* 2004, **16**(4):1021-1034.

109. Tessadori F, van Zanten M, Pavlova P, Clifton R, Pontvianne F, Snoek LB, Millenaar FF, Schulkes RK, van Driel R, Voesenek LA *et al*: **Phytochrome B and histone deacetylase 6 control light-induced chromatin compaction in Arabidopsis thaliana**. *PLoS genetics* 2009, **5**(9):e1000638.

110. Pecinka A, Dinh HQ, Baubec T, Rosa M, Lettner N, Mittelsten Scheid O: **Epigenetic regulation of repetitive elements is attenuated by prolonged heat stress in Arabidopsis**. *The Plant cell* 2010, **22**(9):3118-3129.

CHAPTER TWO

ARABIDOPSIS CROWDED NUCLEI (CRWN) PROTEINS ARE REQUIRED FOR NUCLEAR SIZE CONTROL AND HETEROCHROMATIN ORGANIZATION

Abstract

Background

Plant nuclei superficially resemble animal and fungal nuclei, but the machinery and processes that underlie nuclear organization in these eukaryotic lineages appear to be evolutionarily distinct. Among the candidates for nuclear architectural elements in plants are coiled-coil proteins in the NMCP (Nuclear Matrix Constituent Protein) family. Using genetic and cytological approaches, we dissect the function of the four NMCP family proteins in Arabidopsis encoded by the *CRWN* genes, which were originally named *LINC* (*LITTLE NUCLEI*).

Results

CRWN proteins are essential for viability as evidenced by the inability to recover mutants that have disruptions in all four *CRWN* genes. Mutants deficient in different combinations of the four CRWN paralogs exhibit altered nuclear organization, including reduced nuclear size, aberrant nuclear shape and abnormal spatial organization of constitutive heterochromatin. Our results demonstrate functional diversification among CRWN paralogs; CRWN1 plays the predominant role in control

of nuclear size and shape followed by CRWN4. Proper chromocenter organization is most sensitive to the deficiency of CRWN4. The reduction in nuclear volume in *crwn* mutants in the absence of a commensurate reduction in endoreduplication levels leads to an increase in average nuclear DNA density.

Conclusions

Our findings indicate that CRWN proteins are important architectural components of plant nuclei that play diverse roles in both heterochromatin organization and the control of nuclear morphology.

Background

The cellular components and processes that specify nuclear size, shape and internal organization are poorly understood, particularly in flowering plants. Despite similarities at the gross morphological level among all eukaryotic nuclei, such as a double-membrane boundary perforated with nuclear pores, most of the proteins known to affect nuclear structure in animals are not evolutionarily conserved and are consequently difficult to recognize or absent entirely in plant proteomes [111-113]. These observations indicate that the machinery, and perhaps the principles, specifying nuclear organization in flowering plants are distinct from those operating in animals and represent a convergent evolutionary path to a canonical nuclear organization in eukaryotic cells [114].

We demonstrated previously [78, 115] that two paralogous Arabidopsis coiled-coil proteins, originally named LITTLE NUCLEI 1 and 2 (LINC1 & 2), play important roles in specifying nuclear shape and size. Supporting this conclusion, Sakamoto and Takagi recently reported that disruption of *LINC4*, another of the four paralogous genes in this family, leads to reduced nuclear size and loss of elongated nuclear shape in differentiated cells, mirroring the phenotype of *linc1* mutants [19]. These proteins are closely related to NMCP1, Nuclear Matrix Constituent Protein 1, originally identified as a protein residing on the periphery of carrot nuclei and a component of the salt-resistant nuclear matrix [116]. Although NMCP1 and related proteins are plant-specific and share no significant amino acid similarity to lamins, their tripartite structure with an extensive central coiled-coil domain and their localization at the nuclear periphery suggest that NMCP1-related plant proteins might be functional analogs of this core component of the animal nuclear lamina [76, 77]. More recent computational analysis [80], however, has suggested that the NMCP class of plant proteins shares more structural similarities to myosins or paramyosins than to lamins.

Here, we extend our reverse genetic analysis to encompass all four members of the Arabidopsis NMCP-related protein family, which we have renamed CRWN (CROWDED NUCLEI) to avoid confusion with the acronym LINC (LINKER of NUCLEOSKELETON and CYTOSKELETON) that refers to SUN-KASH protein linkages that bridge the inner and outer nuclear membrane [65, 117-120]. Our findings demonstrate that CRWN proteins are essential for viability, and our analyses uncover complex functional diversification among CRWN proteins with regards to

their effects on whole-plant morphology, nuclear size, and the spatial organization of constitutive heterochromatin aggregates (chromocenters) in interphase nuclei. We found that CRWN1 plays the most prominent role among CRWN paralogs in controlling nuclear size, while CRWN4 has the most important role in controlling the distribution and number of heterochromatic chromocenters. The reduced nuclear size in *crwn* mutants is not matched by a commensurate reduction in endopolyploid levels, resulting in increased nuclear DNA densities (mass per unit volume) up to four-fold higher than wild type levels.

Results

We performed a phylogenetic analysis of Arabidopsis CRWN proteins and their homologues in other species to begin our investigation of the potential diversification within this family. The predicted Arabidopsis proteome contains four closely related CRWN proteins (CRWN1 through 4) that share 30-40% amino acid identity; no other Arabidopsis proteins with extended regions of significant amino acid identity to CRWN proteins were found. Similar proteins were found in other plant species, but interestingly, no fungi or animal CRWN homologues were identified from searches of protein databases. In addition, CRWN homologues were absent in the predicted proteome of the green algae Chlamydomonas and Volvox.

We constructed a phylogram of CRWN proteins and related plant homologs using a maximum likelihood algorithm (Figure 2.1 and Supplementary Table 2.1). The tree

features two major clades distinct from CRWN homologues in two basal plants, *Selaginella moellendorffii* and *Physcomitrella patens*. One clade includes three of the Arabidopsis paralogs, CRWN1, CRWN2 and CRWN3, while CRWN4 belongs to the other clade. Within each clade, the monocot proteins, represented by maize, sorghum and rice, group independently from the dicot proteins. Only two CRWN paralogs exist in these monocots – one CRWN1-like and one CRWN4-like. However, certain dicot species, such as Arabidopsis, poplar, grape, and castor bean, contain multiple copies of CRWN1-like proteins. The dicot CRWN4-like proteins are also distinct from their monocot counterparts in lacking a conserved amino acid motif at the extreme C-terminus (yellow inset in Figure 2.1 and Supplementary Figure 2.1).

Genetic redundancy in the CRWN family

The inference that CRWN4 and related proteins are divergent from members of the CRWN1-containing clade was supported by genetic analyses to dissect the functions of the *CRWN* paralogs. We used Agrobacterium T-DNA insertion alleles to study the effects of inactivating different combinations of *CRWN* genes [121]. Previously, we demonstrated that the *crwn1-1* and *crwn2-1* T-DNA alleles severely reduce or eliminate transcription downstream of the T-DNA insertion [78]. Here, we performed transcript analysis by RT-PCR for the *crwn3-1* and *crwn4-1* alleles used in this study (Supplementary Figure 2.2). For *crwn3-1*, some transcript was detected downstream of the insertion; however, no transcript could be detected using primers that flanked the insertion. The *crwn4-1* insertion blocked transcription downstream of the T-DNA.

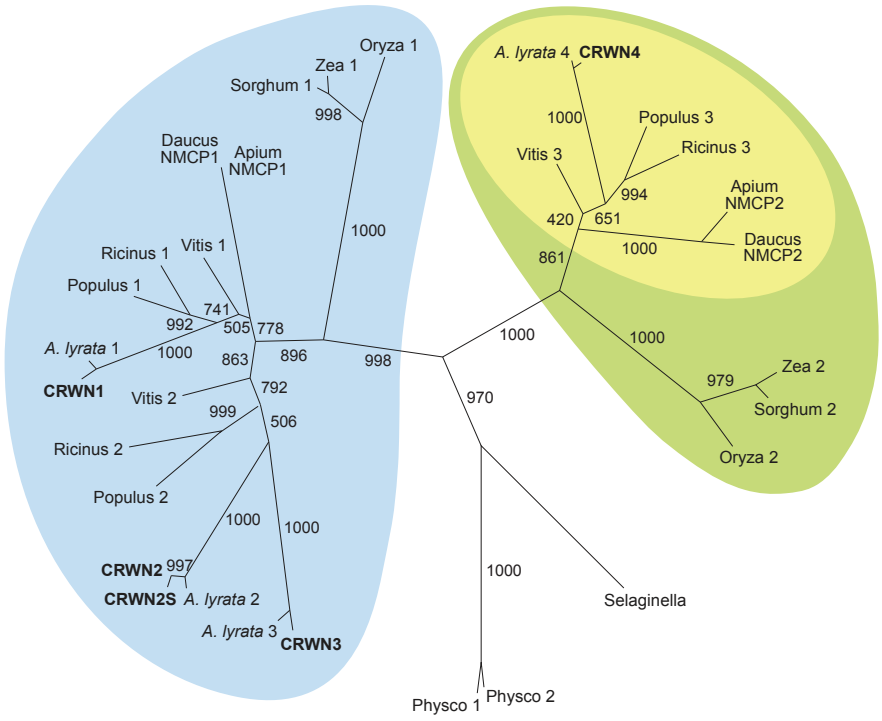


Figure 2.1 Phylogenetic relationships among CRWN proteins.

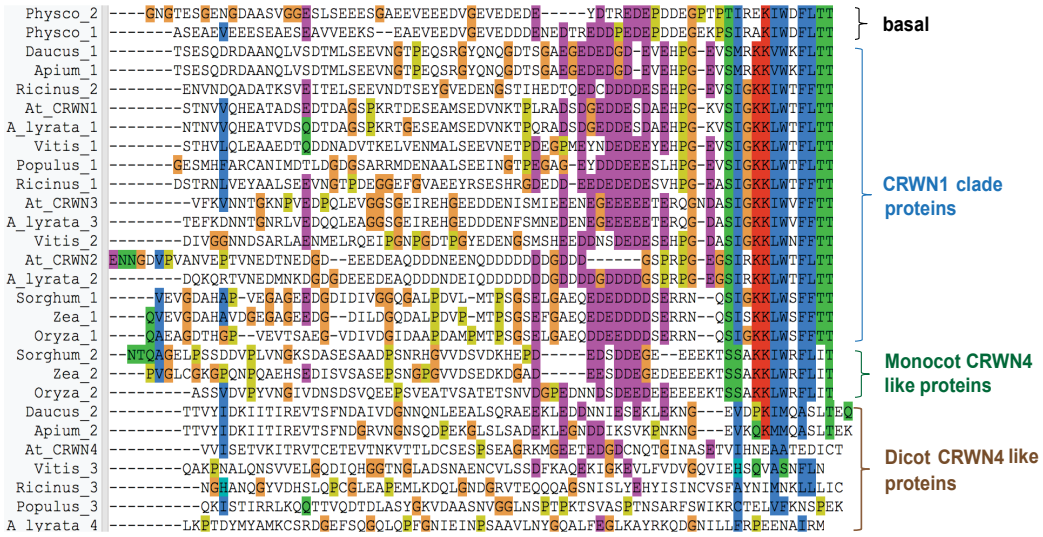
Figure 2.1 A maximum likelihood tree of CRWN homologs constructed from an alignment of amino acid sequences that correspond to the coiled-coil domains. Bootstrap values (of 1000 replicates) are indicated on each branch. The *A. thaliana* CRWN proteins are indicated in bold, and homologs are labeled with the genus name and an assigned number (Supplementary Table 2.1). Two major clades are marked by blue and green; the yellow inset oval indicates a subgroup of CRWN4-like proteins from dicots that lack the conserved C-terminal domain (Supplementary Figure 2.1).

Supplementary Table 2.1 CRWN-like proteins used in this study

Protein	Species name	Locus name or Sequence ID	Database source*
Physco 1	*Physcomitrella patens*	Pp1s200_64V6.1	JGI
Physco 2		Pp1s76_81V6.1	
Selaginella	*Selaginella moellendorffii*	XP_002993584.1	GenBank
Apium NMCP1	*Apium graveolens*	BAF64421.1	GenBank
Apium NMCP2		BAI67716.1	
Daucus NMCP1	*Daucus carota*	BAA20407.1	
Daucus NMCP2		BAI67718.1	
Ricinus 1	*Ricinus communis*	XP_002525969.1	GenBank
Ricinus 2		XP_002524388.1	
Ricinus 3		XP_002530596.1	
Vitis 1	*Vitis vinifera*	CAO49297.1 (GSVIVG01031076001)	GenBank & (JGI)
Vitis 2		CAN74873.1 (GSVIVT01011972001)	
Vitis 3		CAO17747.1 (GSVIVT01007428001)	
Populus 1	*Populus trichocarpa*	XP_002329317.1 (Potri.017G111400.2)	GenBank & (JGI)
Populus 2		XP_002312375.1 (Potri.008G114800.1)	
Populus 3		XP_002317738.1 (Potri.012G034300.1)	
CRWN1	*Arabidopsis thaliana*	At1g67230.1	TAIR
CRWN2		At1g13220.2	
CRWN2S		At1g13220.1	
CRWN3		At1g68790.1	
CRWN4		At5g65770.1	
A_lyrata 1	*Arabidopsis lyrata*	scaffold_2:12,299,939..12,304,385	JGI
A_lyrata 2		471477	JGI
A_lyrata 3		476006	JGI
A_lyrata 4		scaffold_803322.1.1	Ensembl Genomes
Zea 1	*Zea mays*	ZEAMMB73_827243 (AFW63577.1)	Genbank
Zea 2		ZEAMMB73_204423 (DAA57458.1)	
Oryza 1	*Oryza sativa* (Japonica)	Os02g0709900 (NP_001047893.1)	Genbank
Oryza 2		Os01g0767000 (NP_001044359.1)	
Sorghum 1	*Sorghum bicolor*	Sb04g030240.1	JGI
Sorghum 2		Sb03g035670.2	

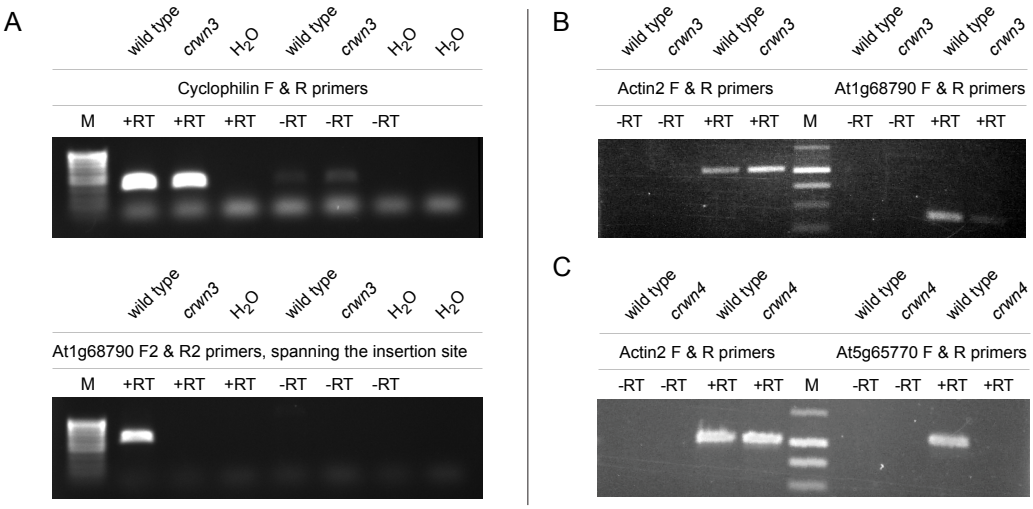
* JGI: www.phytozome.net
GenBank: www.ncbi.nlm.nih.gov/genbank/
TAIR: www.arabidopsis.org
Ensembl Genomes: plants.ensembl.org/Arabidopsis_lyrata/

Supplementary Table 2.1. The first column shows the abbreviated name of the protein used in alignment to construct the tree shown in Figure 2.1. The remaining columns indicate the identity and the source of each protein sequence.



Supplementary Figure 2.1 Amino acid sequences comprising the exreme C termini of 28 CRWN-like proteins, including ten CRWN4-like proteins

Supplementary Figure 2.1 The similarity in this region, which falls outside of the coiled-coil domains, reinforces the topology of the tree shown in Figure 2.1. All of the proteins within the CRWN1-like clade, as well as the Physcomitrella homologs, contain a conserved C-terminal motif and a group of acidic residues approximately 25 amino acids from the end of the protein. Monocot CRWN4-like proteins contain a region with similar features but these conserved motifs are absent in CRWN4-like proteins from dicots (denoted by the yellow oval in Figure 2.1).



Gene/GI	Allele/T-DNA	Insertion Site	RT-PCR Primers	Position of primers
CRWN3 At1g68790	*crwn3-1* SALK_099283	6th exon	F 5'-AGTGAACAGGCAGCTGGTGATAGT-3'	the 6th exon
			R 5'-ACTTCCAACTGCGGATCTTCGACT-3'	the 8th exon
			F2 5'–TCTCCTTCACGGTTTTGAGC–3'	the 6th exon
			R2 5'–GAGAAGCACATGAGGCAGTGT–3'	the 4th exon
CRWN4 At5g65770	*crwn4-1* SALK_079296	6th exon	F 5'-TCGCTAAACCGAGAGCGTGAAGAA-3'	the 6th exon
			R 5'-TTGGTCACCTCTGTCTCACACGTT-3'	the 7th exon
Actin 2			F 5'-TGATATTCAACCAATCGTGTGTGAC-3'	the 1st exon
			R 5'-AAGCAAGAATGGAACCACCGATCC-3'	the 2nd exon
Cyclophilin			F 5'–CGATAAGACTCCCAGGACTGCCGA–3'	the only exon
			R 5'–TCGGCTTTCCAGATGATGATCCAACC–3'	the only exon

Supplementary Figure 2.2 Transcript analysis of the crwn3-1 and crwn4-1 alleles used in this study

Supplementary Figure 2.2 Reverse transcription-PCR results investigating the effect of T-DNA insertions on the transcription of *CRWN3* and *CRWN4*. Panel **A** shows that a *CRWN3* transcript is produced from the wild-type allele but not from the *crwn3-1* allele using primers spanning the T-DNA insertion site. Panel **B** demonstrates that some transcription can be detected downstream of the insertion site from the *crwn3-1* allele using RT-PCR and a primer set recognizing sequences 3' of the insertion site. Panel **C** indicates that the T-DNA insertion in the *crwn4-1* allele blocks transcription. Amplification of cDNA from cyclophilin and *Actin2* were used as positive controls. M, marker lanes; + RT (plus reverse transcriptase); - RT (no reverse transcriptase). Information on the oligonucleotide primers used in these experiments is shown at the bottom of the figure. Our previous data [78] indicated that the *crwn1-1* and *crwn2-1* alleles block transcription downstream of the T-DNA insertion site in the sixth exon of both genes.

The lack of full-length *CRWN* transcripts from homozygous mutant lines indicates that all four mutations used in this study are likely to be loss-of-function alleles. We note that the *CRWN* genes have similar developmental gene expression patterns: the steady-state abundance of transcripts for all four paralogs peak in proliferating tissues (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) [122].

Mutant plants carrying single insertions were intercrossed and progeny carrying homozygous insertions in different combinations were recovered. Figure 2.2 shows the whole-plant phenotype of the viable mutants at the rosette stage just after the transition to flowering. Plants carrying a mutation in any single *CRWN* gene had phenotypes similar to wild-type Columbia plants, as did the double *crwn2 crwn3* and *crwn3 crwn4* mutants. The *crwn2 crwn4* and *crwn1 crwn4* double mutants exhibited slightly smaller rosettes, while the remaining double mutants, *crwn1 crwn2* and *crwn1 crwn3*, displayed markedly smaller rosette sizes. We were able to recover only two of the four triple mutants - *crwn1 crwn2 crwn4* and *crwn1 crwn3 crwn4*, both of which were extremely stunted and set few seed.

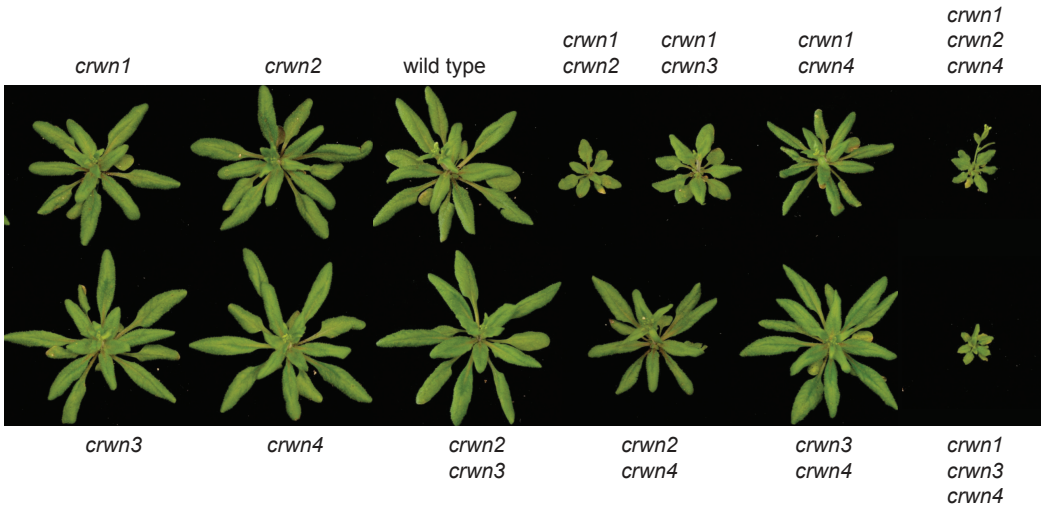


Figure 2.2 Whole plant phenotypes of *crwn* mutants.

Figure 2.2 Leaf rosette structure of one month-old plants imaged just after initiation of flowering. Representative plants for the various genotypes are compared to a wild type (WT) Columbia plant. All plants were grown in parallel and photographed at the same magnification; the diameter of a WT rosette at flowering is ca. 8 cm.

Our inability to isolate a mutant combining alleles in all four *CRWN* genes indicates that at least one functional CRWN protein is required for viability. Triple mutant plants carrying only CRWN2 or CRWN3 were extremely stunted, but still viable. This result suggests that CRWN2 or CRWN3 alone can cover the minimum requirements for the entire CRWN protein family. However, plants carrying only *CRWN1* or only *CRWN4* were not recovered, suggesting that CRWN1 and CRWN4 are specialized and that neither protein alone can express the full range of functions of the CRWN protein family.

CRWN proteins are required to maintain proper nuclear size and shape

We next observed *crwn* mutant nuclei from adult leaf tissue to determine the role of different CRWN proteins in specifying nuclear size and shape. Among the single mutants, a deficiency of *CRWN1* or *CRWN4* reduced nuclear size (Figures 2.3 and 2.4A; Supplementary Table 2.2), while loss of CRWN2 or CRWN3 had no effect. Combining a *crwn1* mutation with a *crwn2* or *crwn3* mutation had a synergistic effect on nuclear size, suggesting that CRWN1 function overlaps, at least partially, with those of CRWN2 and CRWN3. Double mutant combinations containing *crwn4* and either *crwn2* or *crwn3* did not show additive phenotypes but rather resembled *crwn4*. In contrast, combination of a *crwn1* with a *crwn4* mutation had an additive effect on nuclear size. These findings indicate that CRWN1 and CRWN4 are the major determinants of nuclear size among the CRWN paralogs. Further, the additive effects of *crwn1* and *crwn4* mutations suggest CRWN4 acts independently from CRWN1,

consistent with their distinct phylogenetic grouping (Figure 2.1) and the genetic analysis shown in Figure 2.2.

CRWN proteins are required for development or maintenance of the elongated spindle shapes which characterize larger nuclei in differentiated wild-type cells [123]. We previously reported that a deficiency of *CRWN1* causes nuclei in all cells to adopt the spherical shape characteristic of proliferating tissue at root and shoot apices [78]. The present study confirmed the importance of CRWN1 for nuclear shape differentiation and also uncovered a similar role for CRWN4 (Figure 2.3), a conclusion also reached recently by Sakamoto and Takagi [81]. Nuclei from *crwn4* leaf tissue often have irregular margins and are more spherically shaped, compared to wild-type nuclei (Supplementary Figure 2.3). However, *crwn4* nuclei are less uniformly round in comparison to *crwn1* nuclei, particularly larger *crwn4* nuclei. Further, *crwn4* nuclei occasionally contain thin projections that appear to be drawn from the nuclear surface (arrowheads in Figure 2.3).

Supplementary Table 2.2 The nuclear phenotype data for *crwn* mutants used to construct Figure 2.4

	ave. ploidy level	fraction of wt	ave. nuclear size (μm^2)	fraction of wt
wild type	9.74	1.00	73.0 ± 1.18	1.00
crwn1	9.52	0.98	35.7 ± 1.40	0.49
crwn2	8.42	0.86	72.1 ± 2.71	0.99
crwn3	8.41	0.86	69.8 ± 3.43	0.96
crwn4	9.52	0.98	39.3 ± 2.20	0.54
crwn1 crwn2	6.91	0.71	11.8 ± 0.78	0.16
crwn1 crwn3	7.92	0.81	26.5 ± 1.14	0.36
crwn1 crwn4	8.40	0.86	27.8 ± 1.09	0.38
crwn2 crwn3	8.85	0.91	72.9 ± 3.60	1.00
crwn2 crwn4	7.43	0.76	42.5 ± 1.43	0.58
crwn3 crwn4	8.66	0.89	38.1 ± 1.24	0.52
crwn1 crwn2 crwn4	6.31	0.65	10.3 ± 0.31	0.14
crwn1 crwn3 crwn4	5.16	0.59	11.2 ± 0.51	0.15

Supplementary Table 2.2 The nuclear phenotype data for *crwn* mutants used to construct Figure 2.4 is displayed in tabular form. The average endopolyploid level (ave. ploidy level) was determined by flow cytometry as described in Methods. The actual measurements were converted to relative measurements (fraction of wt, third column) using the wild type (wt) values for normalization. The average nuclear size (ave. nuclear size ± standard error of the mean) corresponds to data from Figure 2.4A. The fifth column normalizes these size values to the wild type values.

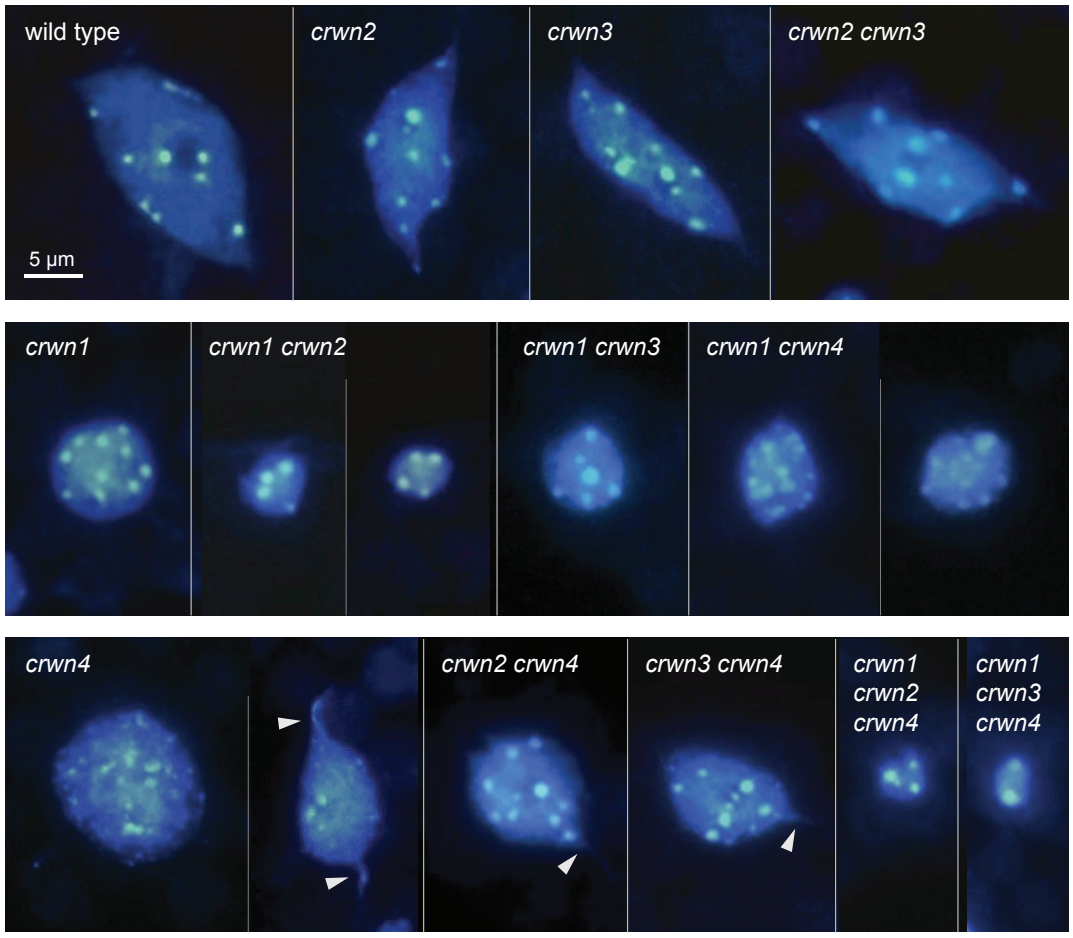


Figure 2.3 Nuclear phenotypes of *crwn* mutants.

Figure 2.3 Images of representative DAPI-stained adult leaf cell nuclei from wild type and the twelve viable *crwn* mutants. The top row contains the wild type control and *crwn2*, *crwn3* and *crwn2 crwn3* mutants with normal nuclear morphology phenotypes. The middle row shows nuclear phenotypes of *crwn1* and double mutants containing a *crwn1* mutation. The bottom row displays nuclear phenotypes from *crwn4* mutants, as well as higher-order mutants containing a *crwn4* mutation. The arrowheads highlight thin projections from *crwn4* nuclei. A 5 µm size bar is shown in the upper left inset; all images are shown at the same magnification.

Loss of CRWN proteins affects nuclear DNA packing density

The direct correlation between endopolyploidy and nuclear size in wild-type Arabidopsis cells [124] prompted us to examine this relationship within the *crwn* mutants. We measured the average endopolyploidy level of nuclei from the same adult leaves harvested for the nuclear size analysis shown in Figure 2.4A (see also Supplementary Table 2.2).

Some *crwn* mutants showed a decrease in endopolyploid levels, particularly the *crwn* triple mutants and the *crwn1 crwn2* double mutant, but the remaining *crwn* genotypes had average endopolyploidy levels that approached wild-type levels (Figure 2.4B). The dashed line in Figure 2.4B depicts the expected nuclear size change in response to a reduction in endopolyploidy based on the established one-to-one relationship between nuclear volume (approximated by nuclear area in our measurements of isolated and flattened leaf cell nuclei, Supplementary Figure 2.4) and DNA content in wild type plants. With the exception of *crwn2* and *crwn3*, the *crwn* mutations caused a more pronounced reduction in nuclear size than predicted from the observed endopolyploidy level. As a consequence, *crwn* mutants display a spectrum of nuclear DNA densities, ranging from wild-type values in *crwn2* and *crwn3* mutants to four-fold higher densities in *crwn1 crwn2* double mutants and the two viable *crwn* triple mutants.

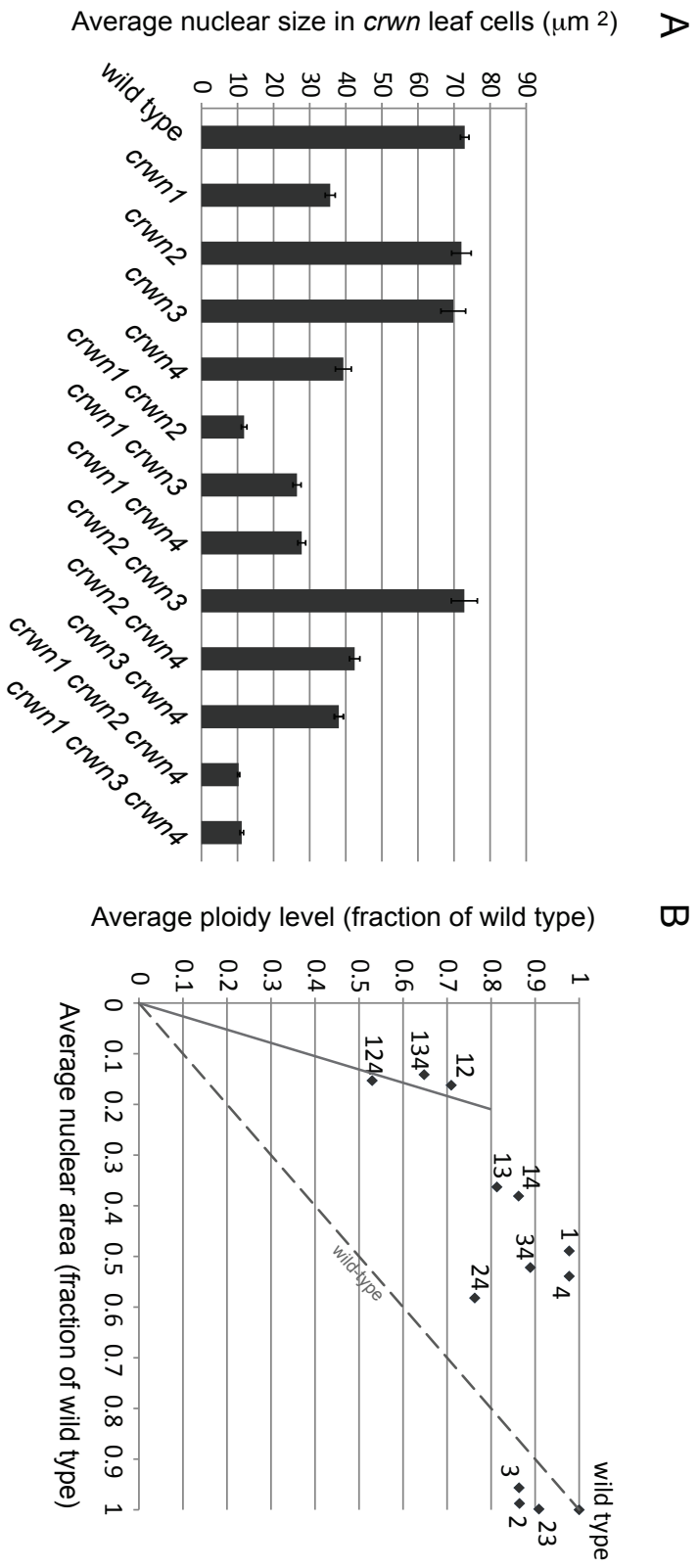
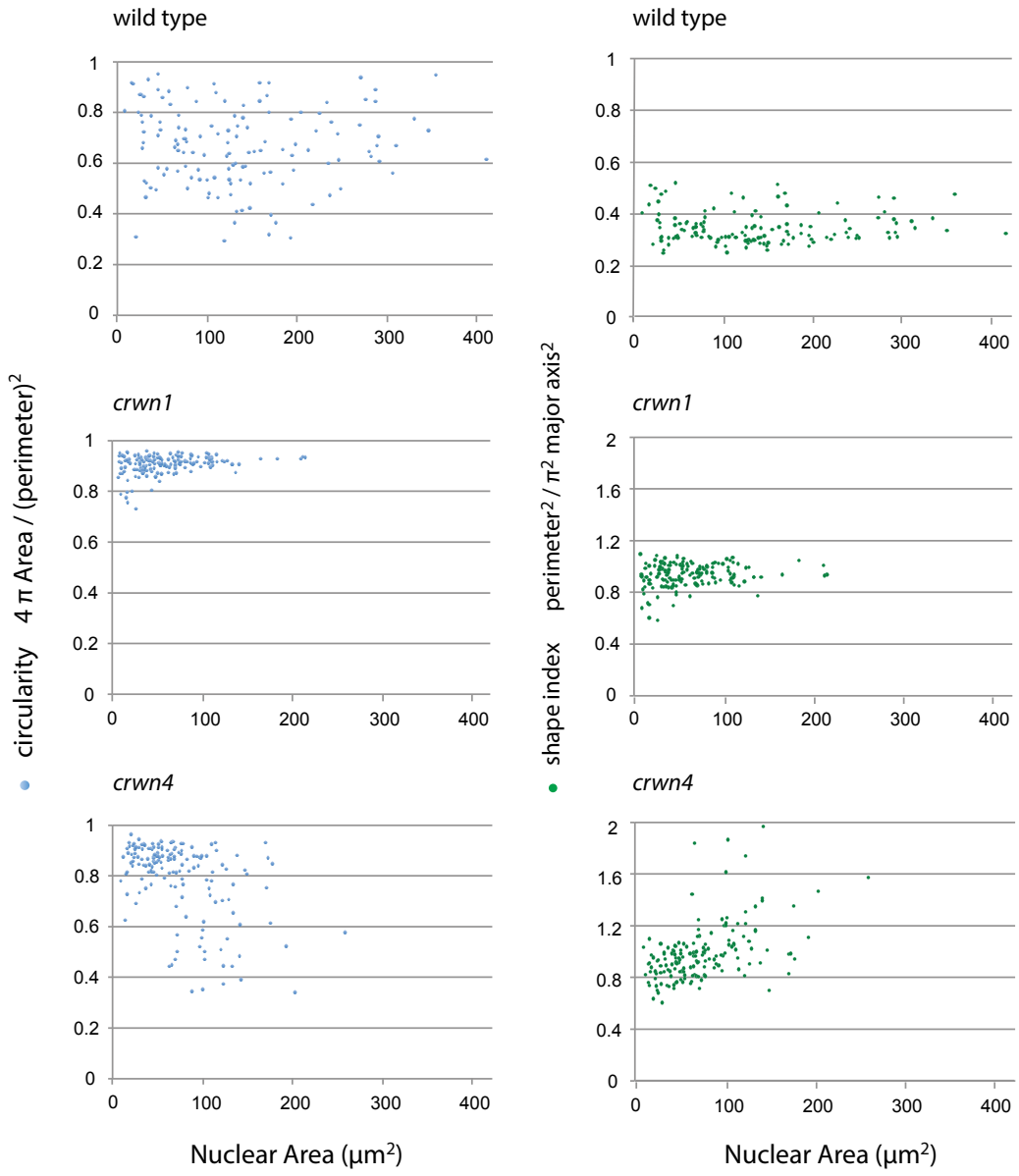


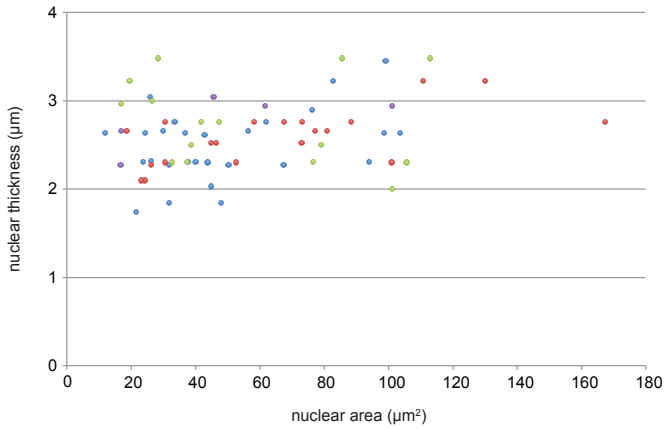
Figure 2.4 The effects of *crwn* mutations on nuclear size and nuclear DNA density in leaf cells.

Figure 2.4 (**A**) Developmentally matched rosette leaves from approximately one month-old plants were fixed, cells isolated and stained with DAPI, and nuclei imaged using epifluorescence microscopy. The average areas of randomly-selected individual nuclei (n = 32-108) were determined for each genotype. Error bars indicated standard error of the mean. (**B**) Nuclear area measurements from panel **A** (see also, Supplementary Table 2.2) were converted to relative values and were plotted against average endopolyploidy level for each genotype expressed as a fraction of the wild-type value. The average endopolyploidy levels (Supplementary Table 2.2) were measured by flow cytometry using corresponding leaf samples. The diagonal dashed line indicates the expected linear relationship between nuclear area and endopolyploid level observed in wild-type plants. Note that panel **A** measures nuclear area but the isolated nuclei are flattened under a coverslip to a uniform thickness (see Supplementary Figure 2.4) and therefore nuclear area is proportional to and therefore a proxy for nuclear volume. The numbers next to the symbols indicate the corresponding *crwn* genotype. Data points above the dashed line indicate an elevated nuclear DNA density relative to wild type. The solid line corresponds to a nuclear DNA density four-times that of wild type.



Supplementary Figure 2.3 Nuclear shape changes in crwn1 and crwn4

Supplementary Figure 2.3 Images of representative DAPI-stained adult leaf cell nuclei from wild type, *crwn1* and *crwn4* mutants were processed by ImageJ software to determine the circularity index ($4\pi \cdot \text{Area} / (\text{perimeter})^2$), as well as a shape index ($\text{perimeter}/\pi \cdot \text{major axis})^2$. Nuclei that deviate from a perfect circle (1.0) show a lower circularity index. The shape index highlights different types of deviations from the round shape. Nuclei in the *crwn1* sample show a shape index close to 1.0, indicating consistently round nuclei. The reduced (relative to 1.0) shape indices in the wild-type sample across all nuclear sizes indicate uniformly elongated nuclear shapes. The elevated shape indices characteristic of larger *crwn4* nuclei result from the presence of thin projections from the surface of otherwise round nuclei.



Supplementary Figure 2.4 Leaf nuclear preparation and confocal imaging reveals a consistent nuclear thickness across a range of nuclear sizes

Supplementary Figure 2.4. Mature leaves were harvested from five individuals in a F2 population segregating both *crwn1* and *crwn2* mutations (F2 plants of a *crwn1 crwn2* x wild type cross). Consequently, the sample captured a range of nuclear shapes and sizes. The nuclei were fixed, isolated, and prepared for imaging as described for Figure 2.4. Following DAPI staining, the three-dimensional signal of different nuclei were recorded and reconstructed using a Leica SP5 confocal microscope. The area of each nucleus was measured using ImageJ, while the thickness of each nucleus was determined by the number and thickness of steps on the z-axis necessary to move from the top to the bottom of each nucleus. The different colored dots on the graph correspond to different slides imaged in this experiment. The results indicate that our preparation and imaging procedure generates nuclei with a relatively uniform thickness, mostly in the 2-3 micrometer range, regardless of the size and shape of the nuclei. Further, this thickness is consistent across individual slides.

We then investigated the relationship between nuclear size and DNA content by examining the effects of different *crwn* genotypes on nuclear size in leaf guard cells, a diploid cell type where endopolyploidy is not a factor [125]. *crwn1* mutant guard cell nuclei were smaller than nuclei in wild type cells with an area approximately one-half of the wild type value, corresponding to a volume difference of approximately threefold assuming a roughly spherical shape to nuclei in the cell (Figure 2.5). Double and triple mutants lacking *CRWN1* displayed nuclear sizes similar to the *crwn1* single mutant. Consistent with their effects on nuclear size shown in Figure 2.4A, neither the *crwn2* nor *crwn3* mutation affected in nuclear size in guard cells. Interestingly, the size of nuclei in *crwn4* guard cells was also unaffected, in contrast to the effect seen in a population of adult leaf cells (Figure 2.4A). However, *crwn2 crwn3*, *crwn2 crwn4*, and *crwn3 crwn4* double mutants had nuclei approximately 20% smaller than those seen in wild-type guard cells, suggesting some functional redundancy among CRWN2, CRWN3 and CRWN4 proteins. Overall, our results indicate that *CRWN1* plays the major role in affecting nuclear size in the absence of changes in endopolyploidy.

CRWN4 maintains interphase chromocenter integrity and organization

Considering the dramatic effects of *crwn* mutations on nuclear size and morphology, we turned our attention to role of CRWN proteins on the internal organization of the nucleus. A conspicuous feature of Arabidopsis interphase nuclei are discrete foci of heterochromatin, or chromocenters, visualized as bright spots after staining with fluorescent DNA-intercalating dyes [126]. A typical interphase nucleus contains

approximately ten chromocenters corresponding to the number of diploid chromosomes (2n=10) [106]. Chromocenter number remains fairly constant over a wide range of nuclear sizes and endopolyploid levels (2n to 16n), most likely as a result of lateral association of sister chromatids after endoreduplication [101, 127]. We found that the average chromocenter number in *crwn1*, *crwn2* and *crwn3* leaf cell nuclei was similar to that seen in wild-type leaf cell nuclei (Figure 2.6A) and did not change dramatically as a function of nuclear size. In *crwn4* nuclei, however, chromocenter number was strongly correlated with nuclear size (Figure 2.6A): smaller nuclei contained fewer chromocenters than the wild-type value of ~9, while larger, presumably endopolyploid, *crwn4* nuclei exhibited a wide range of chromocenter numbers (2-27). A similar pattern was observed in double mutants containing the *crwn4* mutation (Supplementary Figure 2.5). In contrast, double mutants containing the *crwn1* allele paired with another *crwn* mutation displayed a reduced average chromocenter number with a weaker association with nuclear size (Figure 2.6A and Supplementary Figure 2.5).

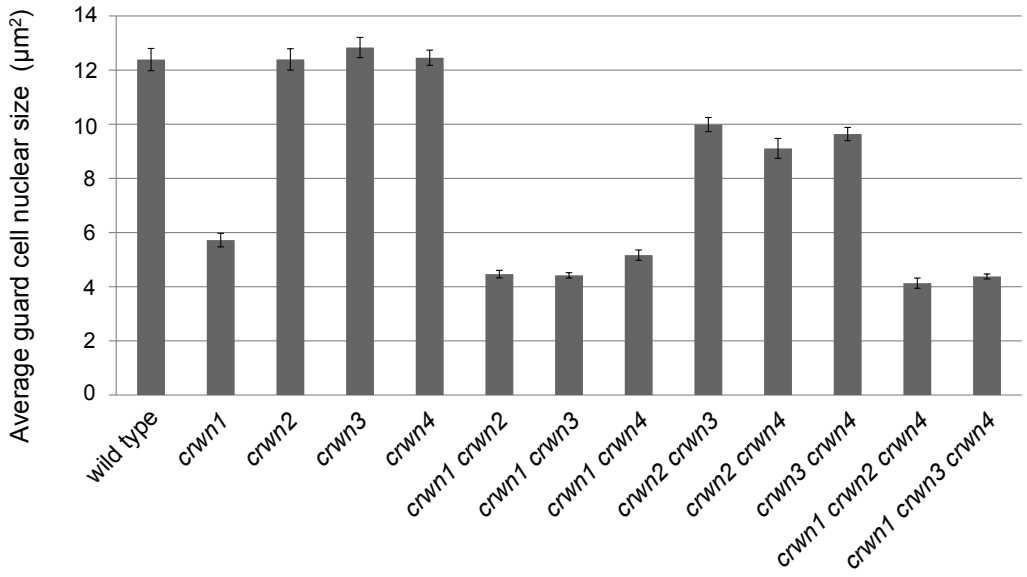


Figure 2.5 Average leaf guard cell nuclear sizes in *crwn* mutants.

Figure 2.5 Two week-old plants were harvested, fixed, stained with DAPI, and leaf guard cell nuclei were imaged using confocal laser scanning microscopy. The area of randomly-selected individual nuclei (n = 19-29) were determined for each genotype. Error bars indicated standard error of the mean.

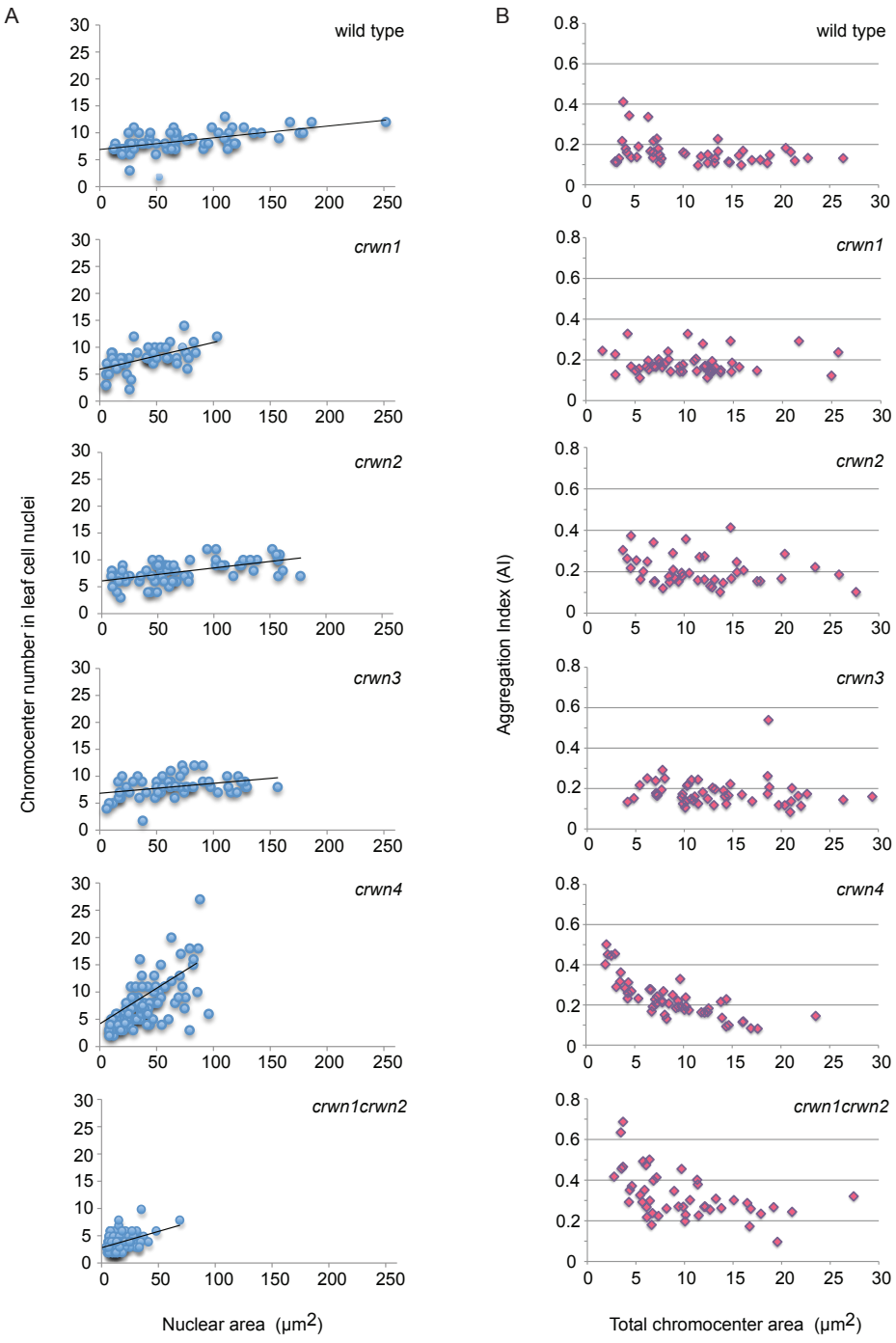
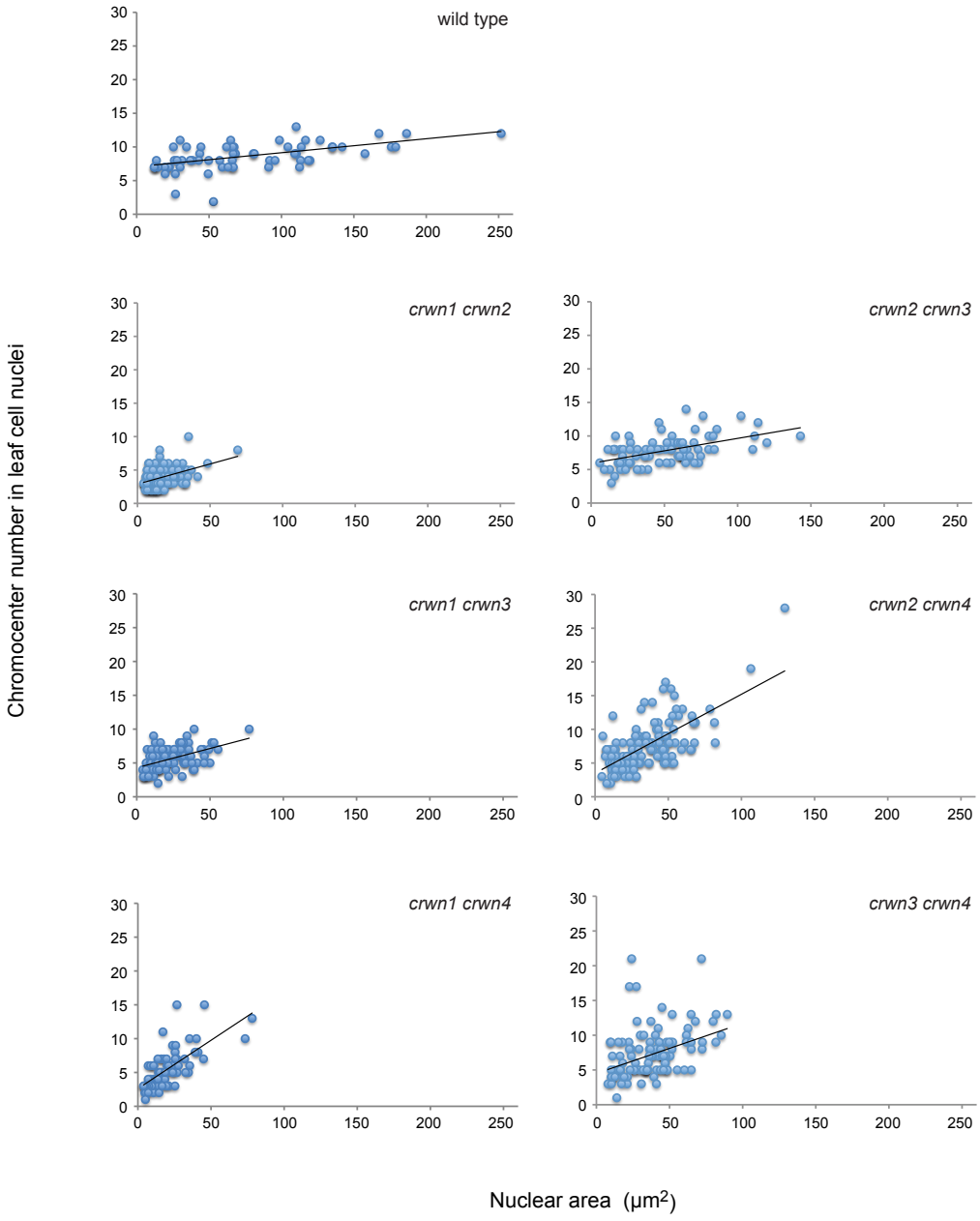


Figure 2.6 Chromocenter morphology changes in *crwn* mutants.

Figure 2.6 (A) Nuclei were imaged from developmentally matched rosette leaves from approximately one month-old plants, cells isolated and stained with DAPI, and nuclei imaged using epifluorescence microscopy. The area and chromocenter number of randomly-selected individual nuclei (n = 47-132) were determined for each genotype, and chromocenter number was plotted against nuclear area. A linear regression line showing the relationship between chromocenter number and nuclear size (as a proxy for endopolyploidy level) was plotted for each genotype. (**B**) Nuclei were imaged fully expanded rosette leaves from approximately one month-old plants and the chromocenter Aggregation Index (AI) was plotted against the total chromocenter area (μm^2).



Supplementary Figure 2.5 Chromocenter changes in crwn double mutants

Supplementary Figure 2.5 Nuclei were harvested from developmentally matched rosette leaves from approximately one month-old plants, stained with DAPI, and imaged using epifluorescence microscopy. The area and chromocenter number of randomly-selected individual nuclei (n = 41-52) were determined for each genotype, and chromocenter number was plotted against nuclear area. A linear regression line showing the relationship between chromocenter number and nuclear size (as a proxy for endopolyploidy level) was plotted for each genotype.

To explore the chromocenter phenotype in more detail, we developed a statistic, referred to as an aggregation index (AI) (see Methods), to characterize the distribution of visible DAPI-bright spots within interphase nuclei. The value of this index ranges from 0 to 1, reflecting both the number of distinct chromocenter spots and the uniformity of their size distribution. The expected AI for wild-type nuclei containing 10 equally sized chromocenters is 0.1, while clustering of chromocenters into fewer but larger aggregates will lead to a higher AI value. A dispersal of chromocenters into smaller heterochromatic satellites will push the AI lower. For a given chromocenter number, a skewed CC size distribution is associated with a larger AI compared to when each CC is equally sized. As shown in Figure 2.6B, the AI index of wild-type nuclei averaged close to 0.1 and was not affected significantly by nuclear size. The absence of a significant correlation between AI and nuclear size indicates that chromocenter organization remains constant across different endopolyploidy levels in wild-type nuclei. A similar pattern was observed for the *crwn1*, *crwn2*, and *crwn3* mutant samples. In contrast, combining *crwn1* and *crwn2* mutations led to an approximately two-fold higher AI over a range of nuclear sizes, consistent with the two-fold reduction in chromocenters number via aggregation in *crwn1 crwn2* mutants. A different pattern was displayed in the *crwn4* sample, which displayed a negative correlation between AI and nuclear size. This result suggests a tendency for chromocenters to aggregate in smaller *crwn4* nuclei and to become dispersed in larger *crwn4* nuclei. The reduction in chromocenter number in *crwn1 crwn2* and *crwn4* mutants with smaller nuclei could reflect the aggregation of individual chromocenters in the limited confines of these nuclei, but a similar clustering does not occur in small

wild-type nuclei, arguing that small nuclear dimensions alone are insufficient to cause clustering. The variability in chromocenter size and number in *crwn* mutant nuclei suggests that CRWN proteins are required for proper organization of heterochromatin in interphase nuclei.

We tested this hypothesis by visualizing the spatial arrangement of chromocenter-associated genomic regions in *crwn1 crwn2* and *crwn4* mutants. Arabidopsis chromocenters are comprised of large segments of repetitive DNA such as the tandemly-arrayed centromeric and 5S RNA repeats located within pericentromeric regions [106]. Using fluorescent *in situ* hybridization (FISH), we examined the spatial organization of the major 180-bp centromeric tandem repeat and the 5S RNA gene arrays in both large and small nuclei from wild-type, *crwn1 crwn2* and *crwn4* plants (Figure 2.7A, B). The centromeric and 5S RNA repeats were co-localized with the DAPI-bright spots in both small and large wild-type nuclei, confirming previous reports that these sequences are normally compartmentalized within chromocenters at the nuclear periphery [128] (Supplementary Movie 2.1 - 2.3, http://www.biomedcentral.com/1471-2229/13/200/additional). It was common to find a decondensed centromere signal at several chromocenters in wild-type nuclei; however, decondensed centromeric repeat clusters were infrequently observed in *crwn1 crwn2* nuclei and the total number of clusters was reduced (Figure 2.7C) (also see Supplementary Movie 2.1 - 2.3). These findings indicate that there is a compaction of the centromere repeat arrays within coalesced chromocenters in *crwn1 crwn2* nuclei. In contrast, the number of discrete centromere repeat clusters visible in

crwn4 nuclei was more variable, and decondensed signals were often seen in nuclei with numerous clusters. This pattern is consistent with the hypothesis that chromocenters become dispersed in larger *crwn4* nuclei. A similar but more pronounced trend was seen for the 5S RNA gene arrays (Figure 2.7B, D), which were dispersed outside chromocenter aggregates in roughly one-half of the *crwn4* nuclei. We note that the dispersed 5S RNA gene signal remained localized to the nuclear periphery (see Supplementary Movie 2.1 - 2.3). The apparent dispersal of chromocenters in larger *crwn4* nuclei and the mis-positioning of centromeric and 5S RNA repeats outside of the chromocenter indicates that higher-order organization of heterochromatin breaks down in interphase in the absence of CRWN4.

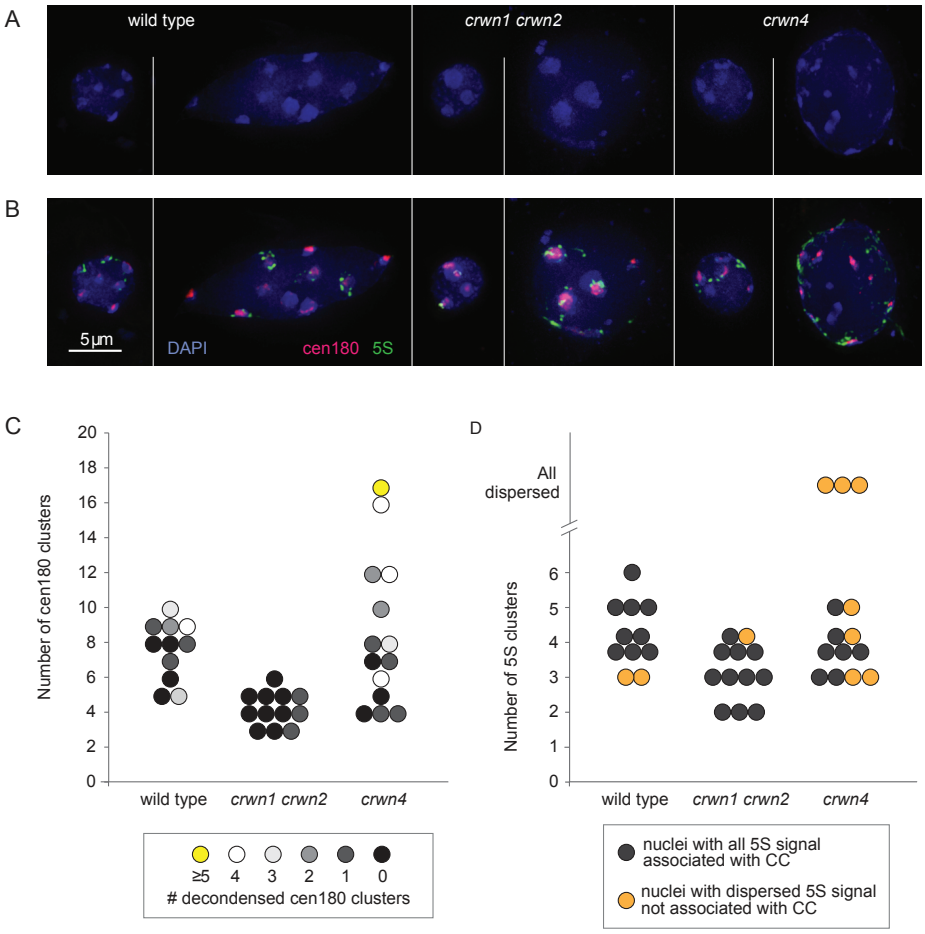


Figure 2.7 Chromocenter organization is altered in *crwn1 crwn2* and *crwn4* mutants. (**A** and **B**) Fluorescence *in situ* hybridization of representative small and large nuclei from wild type, *crwn1 crwn2* and *crwn4* cells prepared from leaves of adult plants after bolting. Blue: DAPI, pink: 180-bp centromere repeats, green: 5S RNA genes, bar, 5 μm. (**C**) The distribution of nuclei based on the number of centromere repeat clusters in different genotypes. Each circle represents an individual nucleus in the FISH experiment. The color indicates the level of decondensation of the centromere signal. (**D**) The distribution of nuclei based on the number of 5S RNA repeat clusters in different genotypes. Each circle represents an individual nucleus. The color indicates the level of dispersion of the 5S repeat signal.

Discussion

Our results demonstrate that the CRWN family is essential for viability in Arabidopsis and required for proper nuclear organization. Redundancy and diversification exists among the four CRWN paralogs, which belong to a plant-specific family of nuclear coiled-coil proteins that forms two clades: one including CRWN1, CRWN2 and CRWN3 and the other containing CRWN4 (Figure 2.1). The divergence of CRWN4 relative to the other CRWN paralogs, also reported by Kimura et al. (2010) [76] and Ciska *et al.* (2013) [77], is supported by our genetic analysis. First, the inviability of triple mutants containing either *CRWN1* or *CRWN4* alone indicates that CRWN1 and CRWN4 possess non-overlapping functions (Figure 2.2). Second, loss-of-function mutations in *CRWN1* and *CRWN4* have distinct phenotypes; for example, *crwn1* is the only *crwn* mutation that affects nuclear size in diploid guard cells (Figure 2.5), while the dispersed chromocenter phenotype is unique to the *crwn4* mutant (Figures 2.3 and 2.7). Finally, the functional distinction between the two CRWN clades is supported by taxa, such as rice and maize, which contain only two CRWN-like proteins: one CRWN1-like and one CRWN4-like.

The phenotypic effects of combining mutations in different *CRWN* genes demonstrates that all four paralogs are involved in specifying nuclear size in adult cells. A deficiency in CRWN1 led to a dramatic reduction in nuclear size independent of an effect on endopolyploidy (Figure 2.4). Loss of CRWN4 also reduced nuclear size in leaf cells (Figure 2.4) as was recently shown by Sakamoto and Takagi (2013) [81].

Interestingly, *crwn4* nuclei in leaf guard cells were not reduced in size (Figure 2.5), despite the fact that *CRWN4* is expressed in this cell type (http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) [122], suggesting cell-type specific requirements exist for different CRWN proteins. Combining a *crwn1* loss-of-function mutation with a deficiency in any of the remaining three paralogs causes a further reduction in nuclear size in leaf cells. The additive effect on nuclear size indicates that CRWN2 and CRWN3 share overlapping functions with CRWN1, although loss of CRWN2 and/or CRWN3 has no effect on nuclear size or endopolyploidy in leaf cells when a functional *CRWN1* allele is present. We previously reported that the *crwn2-1* allele caused a reduction in nuclear size in leaf cells [78]; the reason for the different behavior displayed in this study is unknown, but we have noted variability in nuclear phenotypes among different *crwn2* mutant lines. Regardless, it is clear that CRWN1 plays the major role in adult leaf tissue among the three paralogs in the CRWN1-like clade. This situation might reflect the higher level of expression of *CRWN1* in leaf tissue compared to *CRWN2* and *CRWN3* – ca. 30 FPKM (fragment per kilobase/million mapped RNA-seq reads) for *CRWN1*, with 4x and 2x less expression from *CRWN2* and *CRWN3*, respectively (data not shown). Alternatively, the different contributions of *CRWN1*-like genes could also result from structural differences among the encoded proteins.

Our data indicate that the primary phenotype of *crwn* mutants is a reduction in nuclear size and that the reduction in endopolyploidy observed in double and triple *crwn* mutants is a secondary effect in response to reduced nuclear size. First, mutation of

either *CRWN1* or *CRWN4* had an effect on nuclear size in leaf cells without affecting endopolyploidy (Figure 2.4) (see also [81]). Second, effects on endopolyploid levels were only seen in mutants that contained two or more *crwn* mutations and exhibited severely reduced nuclear size. These considerations indicate that loss of CRWN activity alters the relationship between DNA content and nuclear volume, leading to a higher than normal nuclear DNA density. In the most severely affected mutants, the nuclear DNA density reaches four times the level seen in wild-type cells. It is intriguing that the *crwn* mutants with the most abnormal whole-plant dwarfing phenotypes, including *crwn1 crwn2* and the viable triple *crwn* mutants, are the ones with the highest average nuclear DNA density.

We hypothesize that the CRWN proteins are required for nuclear expansion after nuclear reformation in telophase [42, 129, 130]_ENREF_26, and that the loss of these proteins, especially in combination, results in an elevated nuclear DNA density. Evidence supporting this mechanism comes from a recent report demonstrating that *crwn1 crwn2* cotyledon nuclei expand more slowly than their wild type counterparts during the first 72 hours of seed germination, a period normally characterized by a ca. 10-fold expansion of nuclear size in the absence of endoreduplication [104]. Interestingly, the rate of contraction of *crwn1 crwn2* embryonic cotyledon nuclei during seed maturation is also reduced relative to wild type. These observations suggest that CRWN proteins are involved in remodeling nuclear structure during interphase in response to developmental and environmental cues. Similarly, a

constraint in nuclear expansion associated with endoreduplication [124, 131] could explain the limit on endopolyploid levels observed in high-order *crwn* mutants.

The large coiled-coil domains that comprise the central region of all four CRWN proteins point toward a structural role in the nucleus. Recent analysis based on secondary structure for analogues of Arabidopsis coiled-coil proteins indicates that CRWN proteins are candidates for paramyosin homologues [80]. Paramyosin is a structural protein found in invertebrate muscles, where it forms the core of thick filaments and bundles with myosin motors. Deficiency of paramyosin, encoded by the *unc-15* gene in *C. elegans*, leads to shortened, mis-formed and apparently fragile thick filaments in the nematode's muscles [132]. The structural similarity to paramyosin suggests models for CRWN action as architectural components of the nucleoskeleton. Models of this type predict that *crwn* mutant nuclei would have a less sound structural foundation and be more prone to breakage or distortion by intracellular forces (*e.g.*, exerted by the cytoskeleton or at programmed nuclear expansion/contraction transitions), as seen in animal cells when lamin is disrupted [45] or down-regulated [133]. The irregular margins and thin projections seen among *crwn4* nuclei are consistent with this structural integrity model. These predictions, however, are not consistent with the smaller, round nuclei in *crwn1* mutants that do not adopt the elongated spindle shapes typical of wild-type nuclei in many cell types. These considerations suggest alternative models wherein the CRWN1 protein might establish flexible hinge regions or expansion zones in the nucleoskeleton, facilitating nuclear size changes that accompany endopolyploidy and other developmental transitions. We

note that both nucleoporin 136 [134, 135] and LINC complex mutants [65] in Arabidopsis also lead to more spherical nuclear shapes, suggesting that CRWN proteins might interact with these complexes at the nuclear periphery.

Deficiency of CRWN proteins also affects chromocenter organization (Figure 2.6 and 2.7). We previously reported that chromocenter number decreases approximately two-fold in *crwn1 crwn2* nuclei [78], and we confirmed these results here while extending our analysis to the entire *CRWN* gene family. One unexpected finding was the wide variation in chromocenter number in *crwn4* nuclei and the direct correlation between chromocenter number and nuclear size. The reduction in chromocenter number in *crwn1 crwn2* mutants, as well as *crwn4* mutants with smaller nuclei, is consistent with aggregation of individual chromocenters. The *in situ* hybridization data shown in Figure 2.7 further support this conclusion. Further, we demonstrated that chromocenter organization was disrupted in larger *crwn4* nuclei as evidenced by the dispersed signals seen for the 5S RNA genes, and to a less extent, the centromere repeat arrays. Considering that chromocenters are localized primarily to the nuclear periphery [106, 128] (see Supplementary Movie 2.1 - 2.3) where CRWN proteins are also located [78, 81, 115], CRWN proteins might play a direct role in ensuring proper heterochromatin organization. In such models, CRWN1 and CRWN2 would act to prevent chromocenter aggregation and CRWN4 would exert a complementary effect to maintain chromocenter integrity.

The four distinct phenotypes displayed by *crwn* mutants – reduced nuclear size, altered nuclear shape, elevated nuclear DNA density and abnormal organization of constitutive heterochromatin – highlight a functional connection among these different aspects of nuclear architecture. The lack of whole-plant phenotypes of Arabidopsis *crwn1* and *crwn4* mutants is remarkable in light of the dramatic nuclear changes occurring in these mutants and underscores plants' plasticity in their ability to execute an apparently normal developmental program in spite of these nuclear changes. The diversity of CRWN proteins and the ability to work with viable *crwn* mutants that exhibit dramatic nuclear phenotypes will facilitate the elucidation of the mechanisms through which these essential proteins exert their effects on nuclear organization.

Conclusions

This study addresses fundamental questions about how plant cells specify and control the morphology of their nuclei and its relationship with internal chromatin organization. We conducted a comprehensive reverse genetics study of the *CRWN* gene family in Arabidopsis, which encode NMCP-class proteins implicated in nuclear morphology and organization. We demonstrated that CRWN proteins are essential for viability, and in the process, uncovered a surprisingly high degree of functional diversity among the CRWN proteins. CRWN1 and CRWN4 are the major determinants of nuclear size and shape, and we hypothesize that deficiency in CRWN proteins leads to defects in nuclear expansion and remodeling. One consequence of

this deficiency is an increase in nuclear DNA density as endoreduplication levels are not affected except in the most extreme cases (*e.g.*, *crwn1 crwn2* and the viable triple mutants). Our findings also demonstrated that CRWN4 plays a role in maintenance of heterochromatin organization in interphase nuclei. The specificity of the nuclear morphological and higher-order chromatin organization defects seen in *crwn* mutants reveals the interplay between nuclear morphology and the three-dimensional packaging of the genome.

Materials and Methods

Plant materials and growth conditions

All T-DNA insertion alleles used in this study are from the SALK collection [121] in strain Columbia, and single mutant lines were originally obtained from the Arabidopsis Biological Resource Center (ABRC) at The Ohio State University. Plants were grown in long-day lighting conditions (16 h of light / 8 h of dark) at 23˚C on soil (Metro-Mix 360, SunGro, Vancouver) in environmental growth chambers. Genotyping of individual T-DNA alleles was performed by standard PCR using the following pairs of allele-specific primers: SALK_025347 (*crwn1-1*) , 5'-TGC CTT CTC CTC GCT TTT CAA-3' and 5'-TGC GTG AAT GGG AAA GAA AGT TG-3'; SALK_076653 (*crwn2-1*), 5'-GAA GCT CAT TGC TAG AGA AGG GG-3' and 5'-AAC GCT GAT CGT TCA TGT TCC A-3'; SALK_099283 (*crwn3-1*), 5'-TTC TGC ATC TTG ACA CCA TCC AA-3' and 5'-TCG TCG ACT AGT TAA CAA AAT CA-

3'; SALK_079296 (*crwn4-1*), 5'-CGC AAA GCC TTC GAA GAC AAA-3' and 5'-GCT TCA GCC AGC ATT TCA AGC-3'.

Phylogenetic tree construction

Amino acids similar to CRWN1 were downloaded from public databases (see Supplementary Table 2.1). The program ClustalX was used to align the amino acid sequences. The tree in Figure 2.1 is based on an alignment of the region of highest conservation across all amino acid sequences, corresponding to the coiled-coil domains (amino acids 64 to 651 in CRWN1). A maximum likelihood tree was constructed using Phylip 3.69 with 1000 bootstrap replicates.

Diploid guard cell nuclear area measurement

Two-week old seedlings were harvested and fixed in 3:1 acetic acid:ethanol. Nuclei in the fixed tissue were stained using DAPI (10μg/ml, 2 minutes), and guard cell nuclei were imaged using laser scanning confocal microscopy (Leica SP5). Images were taken at the focal plane with the maximum nuclear area, and the resulting images were processed using ImageJ software.

Leaf nuclei isolation and imaging

Nuclei were isolated from the fifth true leaf of adult plants harvested after initiation of flowering stem elongation (stem height ≥ vegetative rosette diameter). Therefore, the tissues were developmentally matched across the different genotypes. Mesophyll cells predominated but other cell types were present. Each harvested leaf was bisected and one half used for the nuclear area measurement, while the remaining half leaf was processed for flow cytometry measurements (see below). Leaf tissue was fixed using a 3:1 acetic acid:ethanol solution and tissues were rehydrated in 100 mM sodium citrate buffer pH 4.8 for 15 min followed by incubation in digestion buffer (0.03% cytohelicase, 0.03% pectolyase, and 0.03% cellulase Onozuka RS in 100 mM sodium citrate buffer pH 4.8) for 2 hours at 37˚C. Digested tissue was carefully homogenized by pipetting, centrifuged briefly at low speed, and resuspended with 100 mM sodium citrate buffer pH 4.8; this cycle was repeated three times and the final pellet was resuspended in 3:1 acetic acid:ethanol. The resulting suspension of nuclei were pipetted onto microscope slides, dried for ca. 1 min, and stained with 10µg/ml 4',6-diamidino-2-phenylindole (DAPI). A Leica DM 5500 epifluorescence microscope was used to image the nuclei, and the nuclear area was measured from digital images using ImageJ software after manual tracing of nuclear boundaries. Note that the chromocenter number versus leaf cell nuclear area scatter plots shown in Figure 2.6A were generated in a separate experiment from the one shown in Figure 2.4, but developmentally matched leaf tissues were harvested as described above.

Aggregation Index measurement

Tissue from adult leaves was harvested and prepared as described for FISH (see below). Rather than performing the *in situ* hybridization step, nuclei were stained with DAPI and imaged using optical sectioning microcopy. Projections of the processed images were analyzed using ImageJ software to identify chromocenters. Briefly, the images were manually manipulated when necessary to adjust the local threshold and chromocenter area and number were assigned using the Analyze Particle function of the software. The aggregation index (AI) was calculated using the following equation: $AI = \Sigma (S_i/S_{total})^2$; where i = 1, … , n; S_i = the area of chromocenter i; S_{total} = the total area of all chromocenters in the nucleus.

Flow cytometry

Bisected tissue from the fifth true leaf of adult plants were harvested (see above) and immersed in magnesium sulfate buffer [136] and chopped with razor blades in a petri dish. The resulting suspension was filtered through a nylon mesh (diameter = 30 μm; Partec Cell Trics ®, Münster, Germany). The nuclear suspension was incubated with RNAse A (Ribonuclease A, from bovine pancreas, Sigma, St. Louis, MO, USA) (50 μg/ml) on ice for 15 min and stained with propidium iodide (50 μl/ml) in the dark for 6 hours. Average ploidy level for each genotype was calculated based on the peaks generated from an analytical flow cytometer (Accuri 6 model, Accuri cytometers, Ann Arbor, MI, USA).

Fluorescent in situ hybridization (FISH)

Fully expanded adult leaf tissue was harvested less than a week after flowering stem elongation, and fixed in Buffer A [137] with 4% formaldehyde at room temperature with agitation for >1 hour. After rinsing with Buffer A, the tissue was chopped repeatedly with razor blades until a homogenous texture was achieved. A clear nuclear suspension was pipetted from the leaf debris and used for fluorescence *in situ* hybridization as described by Golubovskaya *et al.* (2002) [137]. The centromere probe 5'- Cy5-GGTTGCGGTTTAAGTTCTTATACTCAATC -3' was synthesized by Integrated DNA Technologies (Coralsville, IA, USA), and the 5S probe was amplified from genomic DNA using primers 5'-CTNCCNGGNAGNTCACCC-3' and 5'-CCTNGTGNTGNANCCCTC-3', followed by labeling using a nick translation protocol and Rhodamine labeled dCTP (Roche; Indianapolis, IN, USA).

REFERENCES

1. Meier I: **The plant nuclear envelope**. *Cell Mol Life Sci* 2001, **58**(12-13):1774-1780.

2. Gruenbaum Y, Goldman RD, Meyuhas R, Mills E, Margalit A, Fridkin A, Dayani Y, Prokocimer M, Enosh A: **The nuclear lamina and its functions in the nucleus**. *Int Rev Cytol* 2003, **226**:1-62.

3. Graumann K, Evans DE: **The plant nuclear envelope in focus**. *Biochem Soc Trans* 2010, **38**(Pt 1):307-311.

4. Rose A: **Open Mitosis: Nuclear Envelope Dynamics**. In: *Cell Division Control in Plants*. Edited by Verma DPS, Hong Z. Berlin Heidelberg: Springer-Verlag; 2007.

5. Dittmer TA, Stacey NJ, Sugimoto-Shirasu K, Richards EJ: **LITTLE NUCLEI genes affecting nuclear morphology in *Arabidopsis thaliana***. *Plant Cell* 2007, **19**(9):2793-2803.

6. Dittmer TA, Richards EJ: **Role of LINC proteins in plant nuclear morphology**. *Plant Signal Behav* 2008, **3**(7):485-487.

7. Masuda K, Xu ZJ, Takahashi S, Ito A, Ono M, Nomura K, Inoue M: **Peripheral framework of carrot cell nucleus contains a novel protein predicted to exhibit a long alpha-helical domain**. *Experimental Cell Research* 1997, **232**(1):173-181.

8. Ciska M, Masuda K, Moreno Diaz de la Espina S: **Lamin-like analogues in plants: the characterization of NMCP1 in *Allium cepa***. *J Exp Bot* 2013, **64**(6):1553-1564.

9. Kimura Y, Kuroda C, Masuda K: **Differential nuclear envelope assembly at the end of mitosis in suspension-cultured *Apium graveolens* cells**. *Chromosoma* 2010, **119**(2):195-204.

10. Gardiner J, Overall R, Marc J: **Putative Arabidopsis homologues of metazoan coiled-coil cytoskeletal proteins**. *Cell Biol Int* 2011, **35**(8):767-774.

11. Razafsky D, Hodzic D: **Bringing KASH under the SUN: the many faces of nucleo-cytoskeletal connections**. *J Cell Biol* 2009, **186**(4):461-472.

12. Sosa BA, Rothballer A, Kutay U, Schwartz TU: **LINC complexes form by binding of three KASH peptides to domain interfaces of trimeric SUN proteins**. *Cell* 2012, **149**(5):1035-1047.

13. Tapley EC, Starr DA: **Connecting the nucleus to the cytoskeleton by SUN-KASH bridges across the nuclear envelope**. *Curr Opin Cell Biol* 2013, **25**(1):57-62.

14. Zhou X, Graumann K, Evans DE, Meier I: **Novel plant SUN-KASH bridges are involved in RanGAP anchoring and nuclear shape determination**. *J Cell Biol* 2012, **196**(2):203-211.

15. Graumann K, Evans DE: **Plant SUN domain proteins: components of putative plant LINC complexes?** *Plant Signal Behav* 2010, **5**(2):154-156.

16. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R *et al*: **Genome-wide insertional mutagenesis of *Arabidopsis thaliana***. *Science* 2003, **301**(5633):653-657.

17. Winter D, Vinegar B, Nahal H, Ammar R, Wilson GV, Provart NJ: **An "Electronic Fluorescent Pictograph" browser for exploring and analyzing large-scale biological data sets**. *PLoS One* 2007, **2**(8):e718.

18. Chytilova E, Macas J, Sliwinska E, Rafelski SM, Lambert GM, Galbraith DW: **Nuclear dynamics in *Arabidopsis thaliana***. *Mol Biol Cell* 2000, **11**(8):2733-2741.

19. Sakamoto Y, Takagi S: **LITTLE NUCLEI 1 and 4 regulate nuclear morphology in *Arabidopsis thaliana***. *Plant Cell Physiol* 2013, **54**(4):622-633.

20. Jovtchev G, Schubert V, Meister A, Barow M, Schubert I: **Nuclear DNA content and nuclear and cell volume are positively correlated in angiosperms**. *Cytogenet Genome Res* 2006, **114**(1):77-82.

21. Melaragno JE, Mehrotra B, Coleman AW: **Relationship between endopolyploidy and cell size in epidermal tissue of Arabidopsis**. *Plant Cell* 1993, **5**(11):1661-1668.

22. Fransz P, Soppe W, Schubert I: **Heterochromatin in interphase nuclei of *Arabidopsis thaliana***. *Chromosome Res* 2003, **11**(3):227-240.

23. Fransz P, De Jong JH, Lysak M, Castiglione MR, Schubert I: **Interphase chromosomes in Arabidopsis are organized as well defined chromocenters from which euchromatin loops emanate**. *Proc Natl Acad Sci U S A* 2002, **99**(22):14584-14589.

24. Schubert V, Berr A, Meister A: **Interphase chromatin organisation in Arabidopsis nuclei: constraints versus randomness**. *Chromosoma* 2012, **121**(4):369-387.

25. Schubert V, Klatte M, Pecinka A, Meister A, Jasencakova Z, Schubert I: **Sister chromatids are often incompletely aligned in meristematic and endopolyploid interphase nuclei of *Arabidopsis thaliana***. *Genetics* 2006, **172**(1):467-475.

26. Fang Y, Spector DL: **Centromere positioning and dynamics in living Arabidopsis plants**. *Mol Biol Cell* 2005, **16**(12):5710-5718.

27. Bradley MV: **Cell and nuclear size in relation to polysomaty and the nuclear cycle**. *American Journal of Botany* 1954, **41**(5):398-402.

28. Burke B: **The nuclear envelope: filling in gaps**. *Nat Cell Biol* 2001, **3**(12):E273-274.

29. Levy DL, Heald R: **Nuclear size is regulated by importin alpha and Ntf2 in Xenopus**. *Cell* 2010, **143**(2):288-298.

30. van Zanten M, Koini MA, Geyer R, Liu Y, Brambilla V, Bartels D, Koornneef M, Fransz P, Soppe WJ: **Seed maturation in *Arabidopsis thaliana* is**

characterized by nuclear size reduction and increased chromatin condensation. *Proc Natl Acad Sci U S A* 2011, **108**(50):20219-20224.

31. Sugimoto-Shirasu K, Roberts K: **"Big it up": endoreduplication and cell-size control in plants**. *Curr Opin Plant Biol* 2003, **6**(6):544-553.

32. Epstein HF, Casey DL, Ortiz I: **Myosin and paramyosin of *Caenorhabditis elegans* embryos assemble into nascent structures distinct from thick filaments and multi-filament assemblages**. *J Cell Biol* 1993, **122**(4):845-858.

33. Dechat T, Pfleghaar K, Sengupta K, Shimi T, Shumaker DK, Solimando L, Goldman RD: **Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin**. *Genes Dev* 2008, **22**(7):832-853.

34. Friedl P, Wolf K, Lammerding J: **Nuclear mechanics during cell migration**. *Curr Opin Cell Biol* 2011, **23**(1):55-64.

35. Tamura K, Fukao Y, Iwamoto M, Haraguchi T, Hara-Nishimura I: **Identification and characterization of nuclear pore complex components in Arabidopsis thaliana**. *Plant Cell* 2010, **22**(12):4084-4097.

36. Tamura K, Hara-Nishimura I: **Involvement of the nuclear pore complex in morphology of the plant nucleus**. *Nucleus* 2011, **2**(3):168-172.

37. Arumuganathan K, Earle ED: **Estimation of nuclear DNA content of plants by flow cytometry**. *Plant Molecular Biology Reporter* 1991, **9**:229-233.

38. Golubovskaya IN, Harper LC, Pawlowski WP, Schichnes D, Cande WZ: **The *pam1* gene is required for meiotic bouquet formation and efficient**

homologous synapsis in maize (*Zea mays* L.). *Genetics* 2002, **162**(4):1979-1993.

CHAPTER THREE

THE LOSS OF CRWN PROTEINS LEAD TO BROAD TRANSCRIPTIONAL MIS-REGULATION

Abstract

I describe significant transcriptional changes in Arabdiopsis *crwn* mutants, which have reduced nuclear size and altered nuclear shape. The comparison among *crwn* transcriptomic changes confirmed the synergistic interaction between *crwn1* and *crwn2* mutations evident in our previous genetic studies. In addition, my transcriptomic data uncovered a functional suppression between the *crwn1* and *crwn4* mutations, and demonstrated that *CRWN1-like* genes and *CRWN4* regulate the transcription of an overlapping set of genomic loci. Many mis-expressed loci were identified as candidates for mediating the variety of phenotypic changes observed in *crwn* mutants. The mis-regulated transcriptomes of *crwn* mutants feature activation of stress response pathways and altered expression of many nuclear proteins. I propose that the loss of CRWN proteins affects the structure of the nucleoskeleton and/or the organization of the transcriptional machinery, both of which could lead to alteration of transcriptional profiles.

Introduction

The three-dimensional spatial organization of the eukaryotic genome in the nucleus is thought to play an important role in control of gene expression. For example, actively expressed genes are frequently positioned in the vicinity of nuclear pore complexes [138]. Another demonstration of epigenetic regulation of gene expression through spatial contextualization comes from reports of gene mis-expression after relocalization to the nuclear periphery via association with lamin proteins [139].

These experiments have been carried out in diverse eukaryotic models ranging from mammalian cells to single-cell fungi, arguing that a functional connection between nuclear organization and gene expression is a general principle in eukaryotes. Yet few studies have investigated or focused on this level of epigenetic control in plants. I have investigated a set of mutations in *Arabidopsis thaliana* that alter nuclear shape and size. These mutations disrupt members of a small gene family, called *CRWN* (CROWDED NUCLEI), which encodes four paralogous proteins that contain a large central coiled-coil domain [78, 140]. Mutation of either *CRWN1* or *CRWN4* leads to a reduction in nuclear size and a loss of the elongated nuclear shape characteristic of expanded, differentiated cells in this species [81, 140]. Further, *crwn4* mutations lead to a dispersal of the large heterochromatic 'chromocenter' foci in interphase nuclei [140]. These nuclear morphology defects indicate that *crwn* mutations alter nuclear

architecture, and are therefore useful tools to probe the relationship between nuclear organization and gene expression in plants.

Here I report transcriptional changes in selected *crwn* mutants that advance our understanding of the functional interaction among different CRWN proteins. Our results indicate that i) *crwn4* and *crwn1 crwn2* mutants display the most significant transcriptional mis-regulation; ii) *crwn1* and *crwn4* mutations functionally suppressed each other; and iii) *CRWN1-like* genes and *CRWN4* shared an overlapping set of genomic targets in transcriptional regulation. Many nuclear proteins were mis-regulated in *crwn* mutants, providing explanations for both the altered nuclear organization and nuclear function in these mutants.

Results

To determine if a loss of CRWN proteins causes gene expression changes, I performed mRNA-seq profiling in one-month-old adult leaf samples from six genotypes: wild type*, crwn1, crwn2, crwn4, crwn1 crwn2,* and *crwn1 crwn4*. These genotypes cover the full range of phenotypes characteristic of *crwn* mutants, both in terms of nuclear size and shape, as well as heterochromatin organization (summarized in Table 3.1). Strand-specific cDNA libraries of three biological replicates were constructed for each genotype, and sequence reads were generated using an Illumina Hi-Seq platform. Table 3.2 describes the number of loci identified by reads mapped to the reference

Columbia wild-type genome (annotation version: TAIR10). Approximately 27,000 loci in each genotype were detected, and 73% of these loci have sufficient read numbers to support a statistically-meaningful test. These 'tested' loci included approximately 19,000 genes and 1,000 transposable elements.

Figure 3.1A displays the fold change of the tested loci relative to the corresponding q value statistic. Transposable elements are shown in red and genes in blue. The distribution of data points shifted to different extents in the *crwn* mutants to form a broader base corresponding to mis-regulated genes deviating from the wild-type values with a more robust statistical significance (lower q value). This shift was clearly seen for gene loci (blue dots), but was also evident for transposons (red dots). However, most transposon loci clustered toward the top range of the distributions in Figure 3.1A with higher q values – a situation due in part to the low read numbers associated with these normally transcriptionally quiescent loci. Consistent with this explanation, transposons are over-represented among 'non-tested' loci for which too few reads were recovered to support a robust statistical test for expression changes (see Table 3.1). Nonetheless, transposon mis-regulation in *crwn* mutants was evident when all loci, tested and non-tested, were ordered by their physical position from chromosome 1 to 5 (Figure 3.2). The transposon-rich centromeric regions on each chromosome stood out with a more red color pattern indicating a general trend towards up-regulation. In addition, certain regions in the chromosome arms in the *crwn4*, *crwn1 crwn2,* and *crwn1crwn4* samples also showed mis-regulation.

Table 3.1 Summary of phenotypic changes in *crwn* mutants

	crwn1	*crwn2*	*crwn4*	*crwn1 crwn2*	*crwn1 crwn4*
significantly mis-expressed loci	1327	466	5607	4815	2126
whole plant morphology	normal	normal	normal	dwarf, short internode, smaller leaves, more branches, early flowering, longer life span	slightly dwarf, other changes follow *crwn1 crwn2,* but not as visible or severe
cell size	normal			smaller cells	
nuclear shape	round	wild type	round, with fragile margin, become irregular during harsh	round	round
Ave. leaf nuclear size	50%	75% - 100%	50%	25%	30%
Ave. guard cell nuclear size	50%	100%	80%	50%	50%
Ave. endopoly ploidy level	98%	86%	98%	71%	86%
Sister chromatid cohision and heterochromatin (chromocenters, 5S and 180bp	normal	normal	dispersion	aggregation	aggregation and dispersion

Table 3.1 This table summarizes different phenotypic alterations observed in *crwn* mutants.

Table 3.2 Total number of tested genes and TEs

	crwn1	*crwn2*	*crwn4*	*crwn1crwn2*	*crwn1crwn4*
All loci	26976	27072	27505	27860	27521
Tested loci	19878	20053	20193	20371	19957
TE	973	977	925	1034	941
gene	18905	19076	19268	19337	19016
Significantly mis-regulated loci	1327	466	5607	4815	2126
TE (transposable elements)	12	7	53	77	23
gene	1315	459	5554	4738	2103
Non-tested loci	7098	7019	7312	7489	7564
TE (transposable elements)	2449	2419	2667	2835	2757
gene	4649	4600	4645	4654	4807

Table 3.2 Summary of my mRNA-seq analysis. The total number of loci aligned with reads was counted for each *crwn* mutant (orange) and broken down into statistically tested loci (blue) and non-tested loci (green). Among the tested loci, significantly mis-expressed ones are displayed independently (purple). The number of gene and transposable element targets is indicated for each category.

To determine if this pattern was due to transposon blocks within the chromosome arms, I arranged all loci in sequential order by gene locus ID number followed by transposon locus ID number for each of the five chromosomes (Figure 3.2). The regions corresponding to transposable elements were over-expressed in the *crwn* mutants, particularly *crwn4* and *crwn1 crwn2*. Table 3.1 and Figure 3.1B display significantly up-regulated and down-regulated loci using a q statistic cut-off of 0.01, almost all of which were genes rather than transposons. The number of mis-regulated loci varied across *crwn* genotypes: *crwn4* had over 5500 mis-regulated loci, followed by *crwn1 crwn2* (4800); next in abundance were *crwn1 crwn4* (2300) and *crwn1* (1300), while *crwn2* had no more than 500 mis-expressed loci. As illustrated in Figure 3.1B, there was a skew toward higher fold changes in the up-regulated targets, especially in the *crwn1 crwn2* mutant.

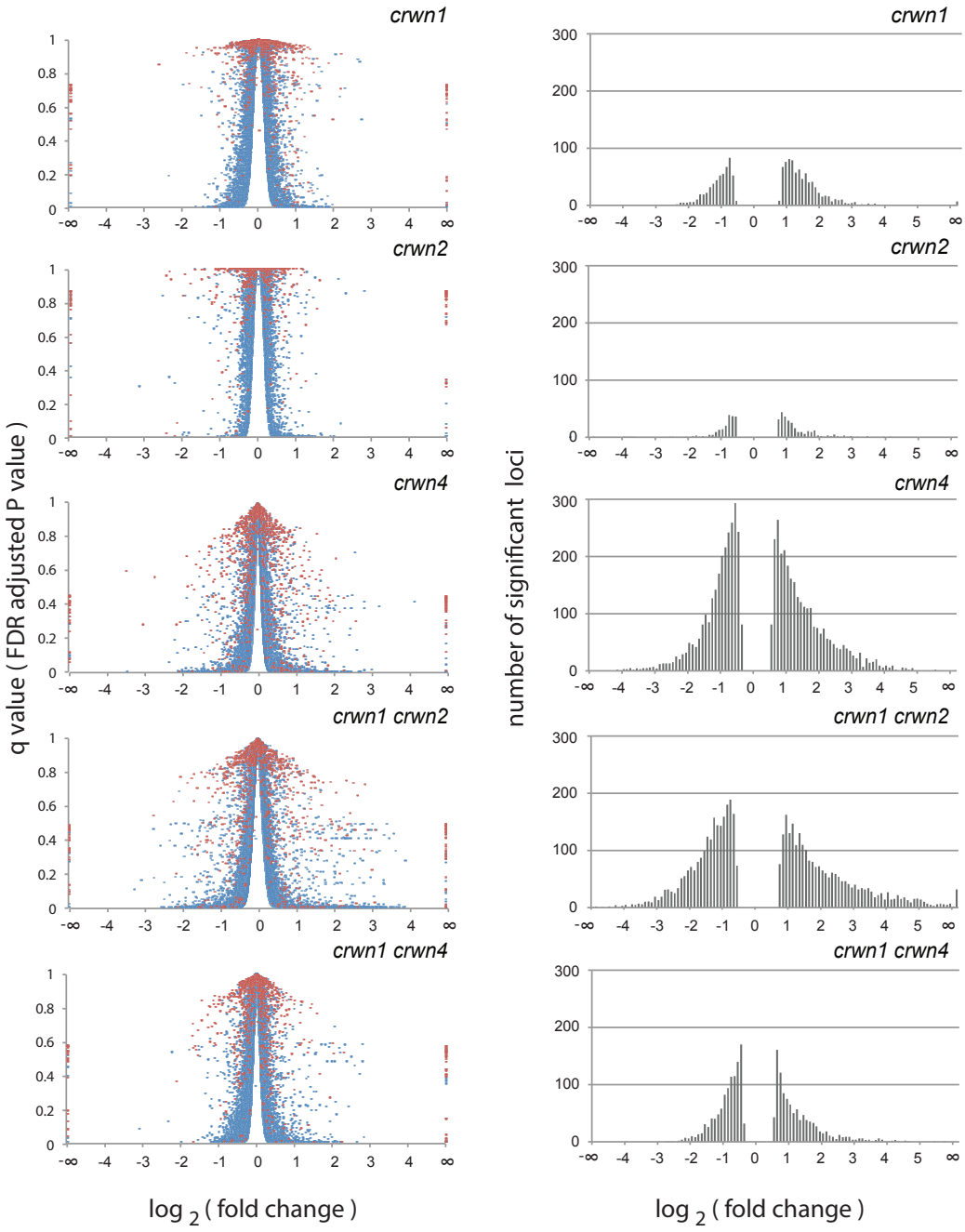


Figure 3.1 The global pattern of transcriptional mis-regulation in *crwn* mutants

Figure 3.1 The global pattern of transcriptional mis-regulation in *crwn* mutants

The scatter plot in panel A (left) displays the $\log_2$ (fold change) value of all tested loci plotted against their corresponding statistic q value for each *crwn* mutant sample. Genes are shown in blue, and transposable elements in red. The histograms in panel B (right) show the distribution of significantly mis-regulated loci in each *crwn* mutant, according to number of loci (Y axis) with $\log_2$ (fold change) value falling into each interval (X axis).

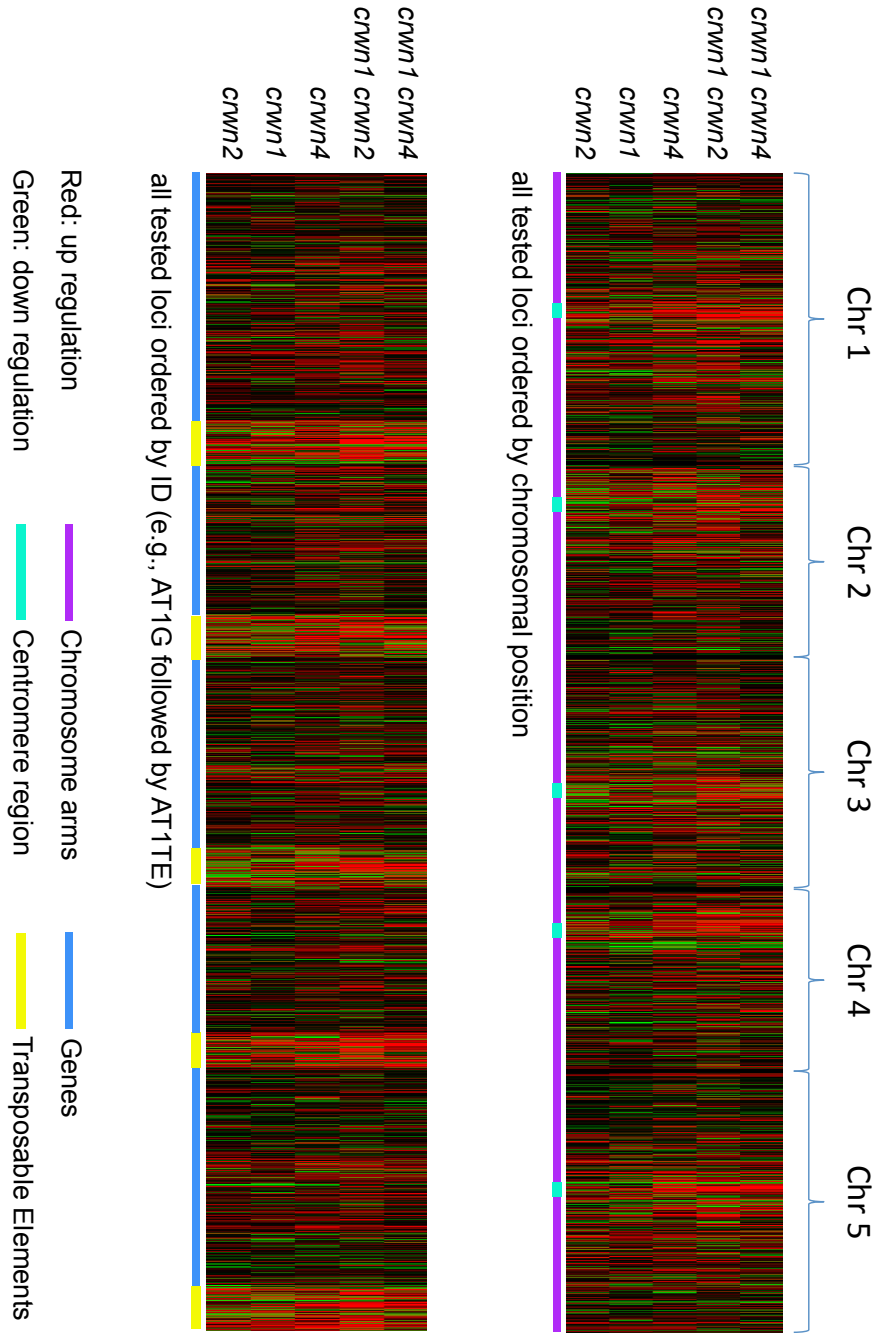


Figure 3.2 Transposon activation in heterochromatic regions

Figure 3.2 The heatmap displays the $\log_2$ (fold change) of all statistically tested loci in five *crwn* mutants in this study. A color gradient from red to green was employed: the extremely up-regulated loci ($+\infty$) were colored in red, while the extremely down-regulated loci ($-\infty$) were colored in green. If no transcriptional change occurs, that locus will be shown in black. In the upper panel, all the loci were ordered by their chromosomal position from left to right. In the bottom panel, the TAIR ID was used as the sorting index; therefore the transposable elements (AtTE#) were distinguished from the genes (AtG#) on each chromosome.

Genetic interactions among the *crwn* mutations

To investigate further the relationship of the altered transcriptomes among different *crwn* mutants, Venn diagrams were used to display the number and identity of the mis-expressed loci from different combinations of multiple *crwn* mutants (Figure 3.3). The up-regulated (left) and down-regulated (right) targets are shown separately, but the patterns of overlap in the left and the right column were consistent for all three combinations. On the top row of Figure 3.3, *crwn1*, *crwn2,* and *crwn1 crwn2* mutants are displayed. *crwn1* and *crwn2* mutants have a relatively small number of significantly mis-regulated loci, represented by small circles, and the majority of *crwn2* mis-regulated targets were shared by the *crwn1* genotype. However, the *crwn1 crwn2* combination led to a much larger number of mis-regulated loci, indicating a synergistic relationship between the effects of *CRWN1* and *CRWN2* deficiencies. This pattern is consistent with our previous results that showed that a combination of *crwn1* and *crwn2* mutations causes a non-additive reduction in nuclear size, as well as an aggregation of heterochromatin and a dwarfing that was absent in either single mutant. Taken together, these data support the conclusion that CRWN1 and CRWN2 possess overlapping functions.

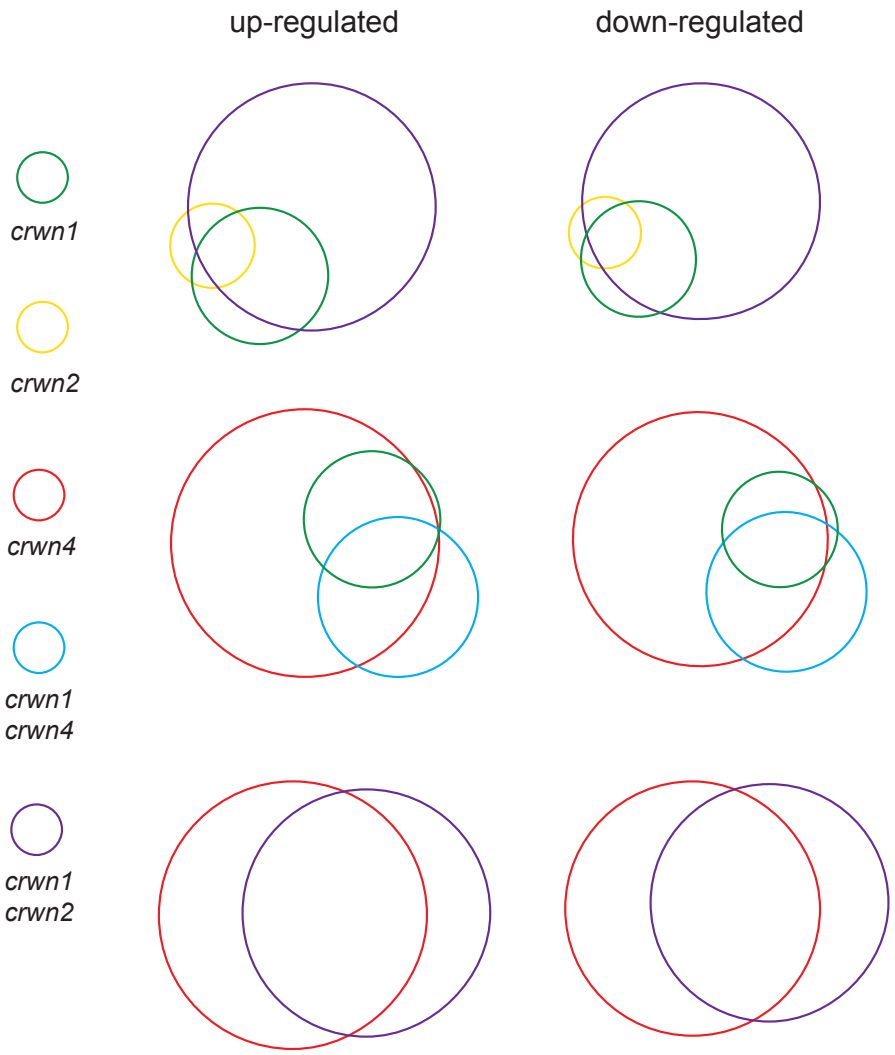


Figure 3.3 The relationship among significantly mis-regulated loci in *crwn* mutants

Figure 3.3 The Venn diagrams illustrate the number of significantly mis-regulated loci among different *crwn* mutants. Each circle with a distinct color represents a particular *crwn* mutant genotype. The size of the circle reflects the number of the loci in each genotype and the overlap between two circles reflects the number of shared mis-regulated loci.

The middle row in Figure 3.3 displays the relationship among mis-regulated loci in *crwn1*, *crwn4,* and *crwn1 crwn4* mutants. As shown in Table 2 and Figure 3.1, the *crwn4* mutant had the largest number of mis-regulated targets among all mutants tested. Interestingly, the majority of *crwn1* mis-regulated loci were also aberrantly expressed in the *crwn4* mutant, indicating that CRWN1 targets a subgroup of genes affected by CRWN4. However, the circle representing *crwn1 crwn4* does not overlap significantly with either the *crwn1* or *crwn4* circle. The number of mis-regulated loci in *crwn1 crwn4* is slightly larger than the number in *crwn1,* but much smaller than that in *crwn4.* This pattern indicates that the *crwn1* mutation partially suppresses the effects of the *crwn4* mutation. This unexpected suppression explains our previous puzzling observation that *crwn1 crwn4* double mutants do not exhibit severe defects on either the nuclear or whole plant level, despite losing the major gene in the CRWN1-like clade and the only gene in the CRWN4 clade.

The bottom row of Figure 3.3 displays the large overlap between *crwn4* and *crwn1 crwn2* mis-regulated targets. Our previous phylogenetic and genetic analyses indicated that CRWN1 and CRWN4 are structurally and functionally diverged. CRWN2 is structurally similar to CRWN1 and these close paralogs possess overlapping functions (see above). The significant overlap of the *crwn1* and *crwn1 crwn2* domains with the *crwn4* circle in the Venn diagram indicates that the apparently specialized CRWN1/2 and CRWN4 functions converge on a common set of transcriptional targets.

To extend the comparison of gene expression patterns among *crwn* mutants to the whole genome, I generated a heatmap that included all statistically tested loci (Figure 3.4). For each locus, a relative expression level among wild type and the five *crwn* mutants was calculated for every genotype, where the denominator was the sum of RPKM (Reads Per Kilobase of exon per Million fragments mapped) values for all genotypes and the numerator is the RPKM value for a specific genotype. The results were sorted by this ratio in the wild type sample and displayed in a color gradient where the lowest ratio is shown in blue and the highest in red. In this heatmap, *crwn1 crwn2* showed the most contrasting color distribution compared to wild type, followed by *crwn4*. The similarity of these patterns indicates that the transcriptome in *crwn1 crwn2* and *crwn4* mutants are disturbed in a similar way. In contrast, *crwn1* had a milder color shift relative to wild type, but the pattern showed some similarity to that of *crwn4*, indicating that *crwn4* and *crwn1* share common targets. The *crwn2* pattern looked most similar to wild type, confirming that the *crwn2* genotype had less transcriptional alteration compared to other *crwn* mutants. These observations are consistent with the genetic relationship among *crwn* mutants revealed by the Venn diagrams shown in Figure 3.3.

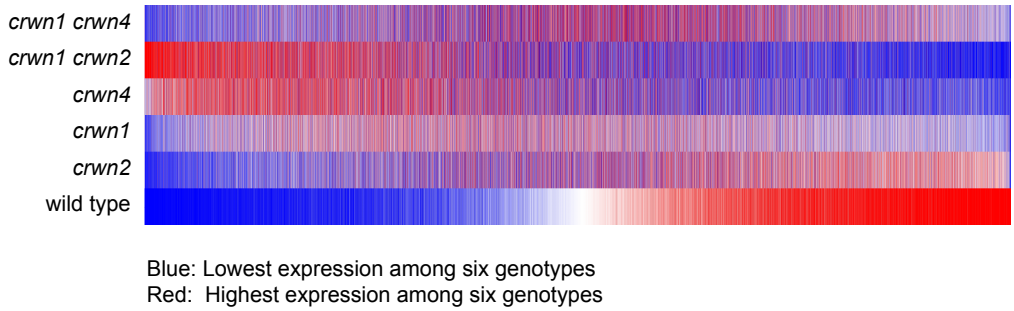


Figure 3.4 Relative expression levels of all statistically tested loci in *crwn* mutants

Figure 3.4 All statistically tested loci from wild type and *crwn* mutants were compared.For each locus, a red-blue color gradient was used to display the relative expression level among all genotypes. The genotype exhibiting the highest column-normalized RPKM value is displayed in a red color, and the lowest in a blue color. Intermediate colors on the red-blue gradient were assigned to other genotypes corresponding to their relative RPKM value at that locus. The arrays of relative expression values were sorted by values for the wild type sample from low (blue) to high (red).

Epigenetic pathways are affected in *crwn* mutants

The observation that many transposons are induced in *crwn* mutants prompted us to investigate whether epigenetic regulatory pathways are affected. I examined the significantly mis-expressed protein coding genes in *crwn* mutants, and determined if there was an enrichment of epigenetically-regulated loci, defined by their mis-expression in *met1, vim1 vim2 vim3, ddc,* and *rdd* mutants where DNA methylation (*met*, *vim*, *ddc*) or de-methylation (*rdd*) pathways were impaired [141]. First, a baseline percentage of the mis-regulated loci in the whole genome was obtained for these epigenetic mutants. The ratio of epigenetically-regulated genes to total mis-expressed loci in *crwn* mutant was then compared to the baseline percentage. Table 3.3 describes the enrichment of these epigenetically-regulated loci in each pathway for *crwn* mutants. Both down-regulated targets of the hypomethylation mutants *met1, ddc,* and *vim1 vim2 vim3*, and the up-regulated targets of the hypermethylation *rdd* mutant are enriched in *crwn* mutant mis-regulated targets. These observations suggest that CRWN proteins are involved in regulating the transcription of epigenetically-regulated loci. Whether the preferentially affected genes are primary or secondary targets, remain unknown.

Table 3.3 Epigenetically controlled loci are affected in *crwn* mutants

TAIR 7 (coding genes)	# of loci	% epi # /	fold crwn % /	number of loci			% epi # / genotype #			fold crwn % / wt %			number of loci			% epi # / genotype #			fold crwn % / wt %		
				up	same	down	up	same	down	up	same	down	up	same	down	up	same	down	up	same	down	
wild type	26994												7	192	109	0.03%	0.71%	0.40%			
met1 up	308	1.14%		7	113	188	0.03%	0.42%	0.70%				3	328	16	0.01%	1.22%	0.06%			
met1 down	347	1.29%		17	208	122	0.06%	0.77%	0.45%				3	253	17	0.01%	0.94%	0.06%	=1	no enrichment	
ddc up	273	1.01%		5	217	51	0.02%	0.80%	0.19%	=1	no enrichment		0	75	8	0.00%	0.28%	0.03%	>1	enriched	
ddc down	83	0.31%		1	67	15	0.00%	0.25%	0.06%	>1	enriched		0	80	0	0.00%	0.30%	0.00%	<1	under represented	
rdd up	80	0.30%		3	66	11	0.01%	0.24%	0.04%	<1	under represented		3	62	1	0.01%	0.23%	0.00%			
rdd down	66	0.24%		7	51	8	0.03%	0.19%	0.03%												
vim1 vim2 vim3 up	104	0.39%																			
vim1 vim2 vim3 down	68	0.25%																			
crwn 4	5409			up	same	down	up	same	down	up	same	down	up	same	down	up	same	down	up	same	down
met1 up	43	0.79%	0.70	4	32	7	0.07%	0.59%	0.13%	2.85	1.41	0.19	0	42	1	0.00%	0.78%	0.02%		1.09	0.05
met1 down	123	2.27%	1.77	8	83	32	0.15%	1.53%	0.59%	2.35	1.99	1.31	0	117	6	0.00%	2.16%	0.11%		1.78	1.87
ddc up	54	1.00%	0.99	1	48	5	0.02%	0.89%	0.09%	1.00	1.10	0.49	0	53	1	0.00%	0.98%	0.02%		1.05	0.29
ddc down	30	0.55%	1.80	1	24	5	0.02%	0.44%	0.09%	4.99	1.79	1.66	0	25	5	0.00%	0.46%	0.09%		1.66	3.12
rdd up	28	0.52%	1.75	0	24	4	0.00%	0.44%	0.07%	0.00	1.81	1.81	0	28	0	0.00%	0.52%	0.00%		1.75	
rdd down	17	0.31%	1.29	3	13	1	0.06%	0.24%	0.02%	2.14	1.27	0.62	1	16	0	0.02%	0.30%	0.00%	1.66	1.29	
vim1 vim2 vim3 up	29	0.54%	1.39																		
vim1 vim2 vim3 down	35	0.65%	2.57																		
crwn1 crwn2	4583			up	same	down	up	same	down	up	same	down	up	same	down	up	same	down	up	same	down
met1 up	52	1.13%	0.99	4	34	14	0.09%	0.74%	0.31%	3.37	1.77	0.44	0	48	4	0.00%	1.05%	0.09%		1.47	0.22
met1 down	94	2.05%	1.60	7	61	26	0.15%	1.33%	0.57%	2.43	1.73	1.26	0	91	3	0.00%	1.99%	0.07%		1.63	1.10
ddc up	53	1.16%	1.14	1	45	7	0.02%	0.98%	0.15%	1.18	1.22	0.81	0	51	2	0.00%	1.11%	0.04%		1.19	0.69
ddc down	40	0.87%	2.84	1	32	7	0.02%	0.70%	0.15%	5.89	2.81	2.75	0	35	5	0.00%	0.76%	0.11%		2.75	3.68
rdd up	22	0.48%	1.62	1	19	2	0.02%	0.41%	0.04%	1.96	1.70	1.07	0	22	0	0.00%	0.48%	0.00%		1.62	
rdd down	15	0.33%	1.34	4	10	1	0.09%	0.22%	0.02%	3.37	1.15	0.74	1	14	0	0.02%	0.31%	0.00%	1.96	1.33	
vim1 vim2 vim3 up	35	0.76%	1.98																		
vim1 vim2 vim3 down	26	0.57%	2.25																		
crwn1 crwn4	2039			up	same	down	up	same	down	up	same	down	up	same	down	up	same	down	up	same	down
met1 up	32	1.57%	1.38	3	17	12	0.15%	0.83%	0.59%	5.67	1.99	0.85	0	29	3	0.00%	1.42%	0.15%		2.00	0.36
met1 down	70	3.43%	2.67	5	41	24	0.25%	2.01%	1.18%	3.89	2.61	2.60	1	68	1	0.05%	3.33%	0.05%	4.41	2.74	0.83
ddc up	38	1.86%	1.84	0	34	4	0.00%	1.67%	0.20%	0.00	2.07	1.04	0	37	1	0.00%	1.81%	0.05%		1.94	0.78
ddc down	24	1.18%	3.83	1	21	2	0.05%	1.03%	0.10%	13.24	4.15	1.77	0	24	0	0.00%	1.18%	0.00%		4.24	
rdd up	14	0.69%	2.32	0	10	4	0.00%	0.49%	0.20%	0.00	2.01	4.81	0	14	0	0.00%	0.69%	0.00%		2.32	
rdd down	11	0.54%	2.21	2	8	1	0.10%	0.39%	0.05%	3.78	2.08	1.65	0	11	0	0.00%	0.54%	0.00%		2.35	
vim1 vim2 vim3 up	19	0.93%	2.42																		
vim1 vim2 vim3 down	14	0.69%	2.73																		
crwn1	1271			up	same	down	up	same	down	up	same	down	up	same	down	up	same	down	up	same	down
met1 up	15	1.18%	1.03	2	8	5	0.16%	0.63%	0.39%	6.07	1.50	0.56	0	14	1	0.00%	1.10%	0.08%		1.55	0.19
met1 down	32	2.52%	1.96	3	18	11	0.24%	1.42%	0.87%	3.75	1.84	1.91	0	28	4	0.00%	2.20%	0.31%		1.81	5.31
ddc up	15	1.18%	1.17	0	12	3	0.00%	0.94%	0.24%	0.00	1.17	1.25	0	15	0	0.00%	1.18%	0.00%		1.26	0.00
ddc down	14	1.10%	3.58	0	13	1	0.00%	1.02%	0.08%	0.00	4.12	1.42	0	12	2	0.00%	0.94%	0.16%		3.40	5.31
rdd up	7	0.55%	1.86	0	7	0	0.00%	0.55%	0.00%	0.00	2.25	0.00	0	7	0	0.00%	0.55%	0.00%		1.86	
rdd down	4	0.31%	1.29	1	3	0	0.08%	0.24%	0.00%	3.03	1.25	0.00	0	4	0	0.00%	0.31%	0.00%		1.37	
vim1 vim2 vim3 up	18	1.42%	3.68																		
vim1 vim2 vim3 down	18	1.42%	5.62																		

Table 3.3 The statistics are shown for the enrichment of epigenetically controlled loci corresponding to significantly mis-expressed loci in *crwn* mutants. The epigenetic pathways examined include: CpG DNA methylation guided by VIM1, VIM2, VIM3 [142, 143] and MET1 [144]; CHG/CHH DNA methylation guided by DRM1, DRM2, and CMT3 [144]; and DNA de methylation guided by ROS1, DML2 and DML3 [144]. In four *crwn* mutants, the total number of mis-regulated genes belonging to *met1, ddc, rdd* or *vim1 vim2 vim3* target lists are displayed in Column B (in blue), the ratios of epigenetically regulated targets over total mis-regulated targets within each genotype were calculated as percentage in Column C (in blue), and the enrichment fold (percentage in *crwn* mutant / percentage in wild type) were shown in Column D (in blue). Genes mis-regulated in *met1*, *ddc* and *rdd* mutants background are further categorized by three classes of alterations in DNA methylation (in orange, Column E - M) or small RNA expression (in purple, Column N – V) at corresponding genomic loci. Similarly, in each class of alteration, the total number, percentage and fold enrichment of epigenetically regulated loci were calculated for these four *crwn* mutants.

Not all epigenetically-regulated targets are affected in *crwn* mutants, however. I broke down the genes affected by the *met1, ddc,* and *rdd* pathway into sub-categories according to whether changes in small RNA accumulation or DNA methylation occurred at each locus. The most commonly enriched epigenetic targets among the mis-expressed loci in *crwn* mutants were the ones with little or no changes in small RNA expression and/or DNA methylation. Interestingly, a major pool of up-regulated targets in the *met1* mutant (109 loci, 35%) included loci for which the transcriptional activation is coupled with reduced small RNA expression, and almost none of these loci are affected by *crwn* mutations. Also, the genes whose up-regulation is associated with CpG hypo-methylation due to the loss of MET1 were not over-represented in *crwn* mutants. A similar bias also applied to primary *ddc* targets, where the CHG or CHH methylation was lost; while an opposite pattern existed among the hypo-methylated *rdd* targets, for which the de-methylation function was impaired. *rdd* mutations also lead to hyper-methylation, and these type of loci, despite their small number, were significantly enriched among mis-regulated loci in all *crwn* mutants. Taken together, this analysis suggests that *crwn* mutations affect a subset of epigenetically-controlled loci.

Categories of loci mis-expressed in *crwn* mutants

Next I explored the functional categorization of genes whose expression was disturbed in *crwn* mutants. Figure 3.5 displays the percentage of significantly mis-regulated loci in different sub-categories (*e.g.*, based on cellular location, molecular function, and

biological process). A genome-wide baseline percentage was calculated for each category using all the annotated loci in TAIR10. I then compared the mis-expressed loci in different *crwn* mutants relative to this default ratio, and found varying levels of under- or over-representation among different categories. Among the different biological processes, many loci involved in response to biotic/abiotic stimuli and stress were significantly over-represented among the genes up-regulated in *crwn* mutants, as were signal transduction and transport loci (Figure 3.5A). Among different molecular functions, transcription factors and kinases were up-regulated (Figure 3.5B). Further, from the perspective of different cellular compartments, the proteins located in plasma membrane were, as a group, up-regulated (Figure 3.5A). These observations are consistent with the activation of signaling transduction proteins at the cell surface that are involved in stress response pathways. In contrast, genes involved in biogenesis were over-represented in the down-regulation group, but under-represented in the up-regulation group. Additionally, proteins targeted to the cell wall and extracellular region were significantly down-regulated. These data suggest that homeostatic functions (*e.g.*, development, biogenesis) are down-regulated in *crwn* mutants, while the defense response cascades are activated.

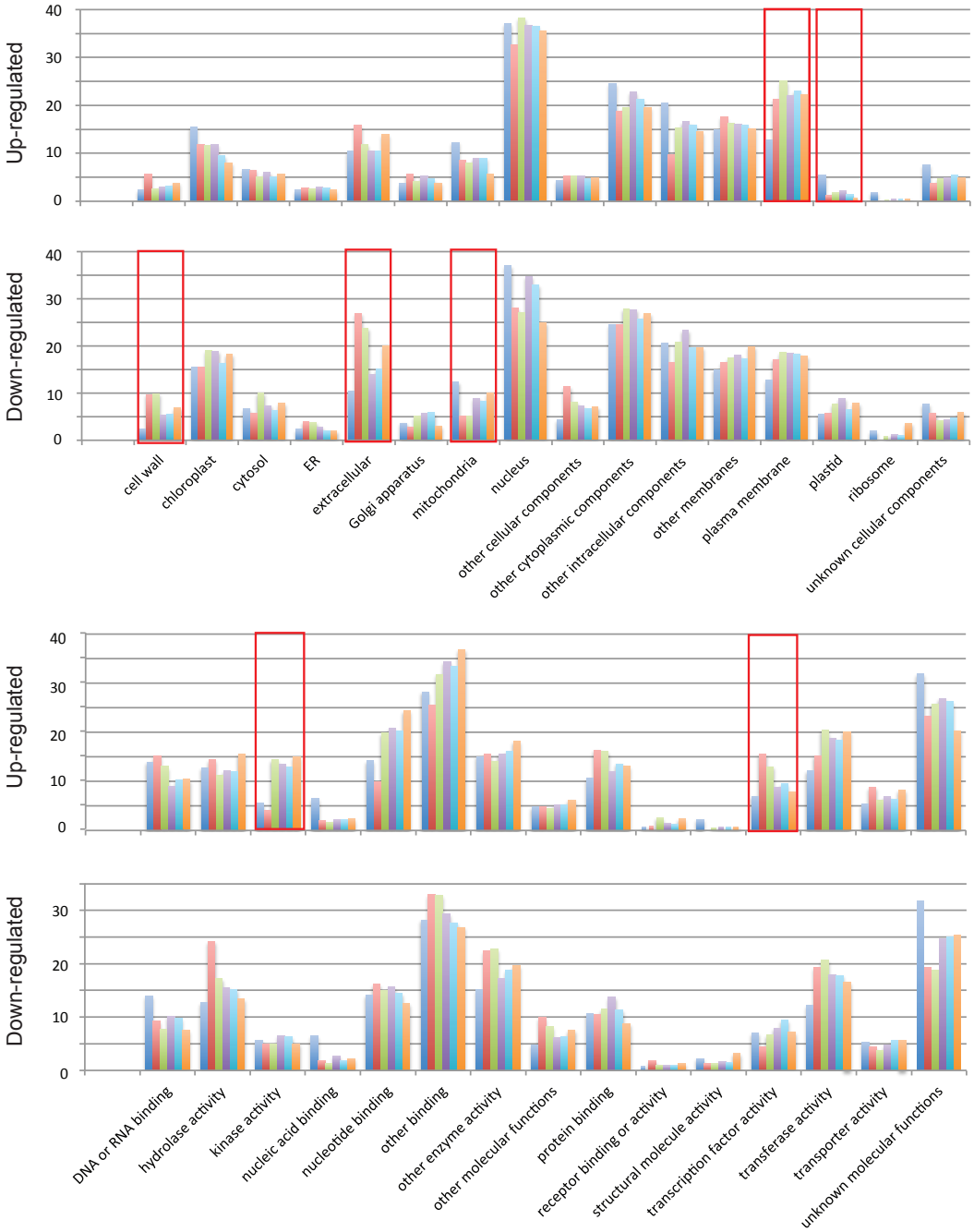


Figure 3.5A Functional categorization of mis-expressed loci in *crwn* mutants

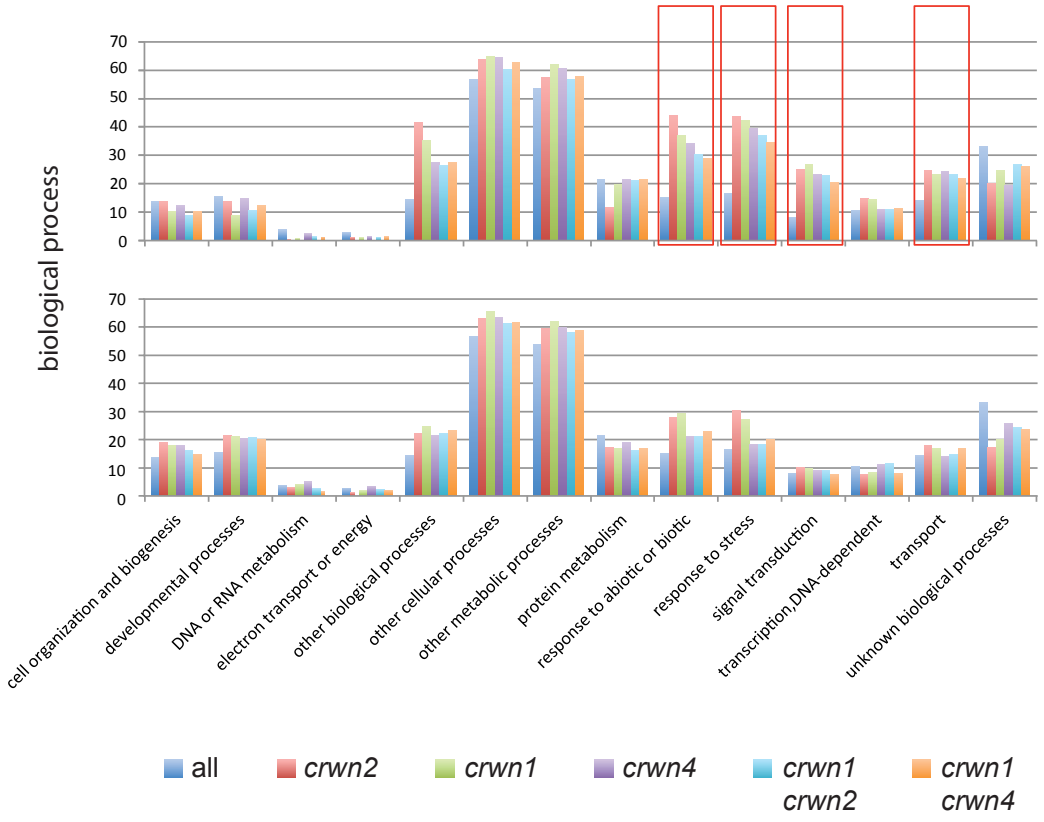


Figure 3.5B Functional categorization of mis-expressed loci in *crwn* mutants

Figure 3.5 Three panels display the functional categorization of genes significantly mis-regulated loci in *crwn* mutants depending on cellular compartment, molecular function, and biological process. In each panel, the up-regulated loci and down-regulated loci are displayed separately. The Y-axis represented the percentage of mis-expressed loci for each sub-category displayed on X-axis. The wild type and the five *crwn* mutant samples are displayed using different colors. Red rectangles were used to highlight those sub-categories that had over- or under-representation of mis-expressed loci among *crwn* mutants in comparison to wild type.

To investigate these provisional conclusions further, I looked more closely at transcription factors that were mis-regulated in *crwn* mutants. The major gene families affected include ERF/AP2, WRKY, NAC, MYB, and bHLH. Among them, the over-representation of ERF/AP2 and WRKY family proteins was the most striking (Figure 3.6A). The Arabidopsis predicted proteome contains 117 ERF/AP2 members [145], belonging to twelve classes (A1-6, B1-6) constituting 7% of all transcription factors. The enrichment of ERF/AP2 proteins was biased to the extremely up-regulated transcription factors in *crwn1* (25%), *crwn1 crwn2* (17%), and *crwn4* (18%), but not in *crwn1 crwn4* (Figure 3.6A). The loci affected encode ERF/AP2 proteins in the A1, A5, B1, and B3 groups predominantly, and involve members that operate in the response to cold, pathogen, and mechanical stresses (Supplementary Table 3.1). The WRKY family consists of 75 members [146], which comprise 4.5% of all the transcription factors genome wide. These WRKY members are enriched among the up-regulated transcription factors in all *crwn* mutants, especially in *crwn1 crwn4* (26%) and *crwn1 crwn2* (17%) (Figure 3.6A; Supplementary Table 3.1).

Figure 6A

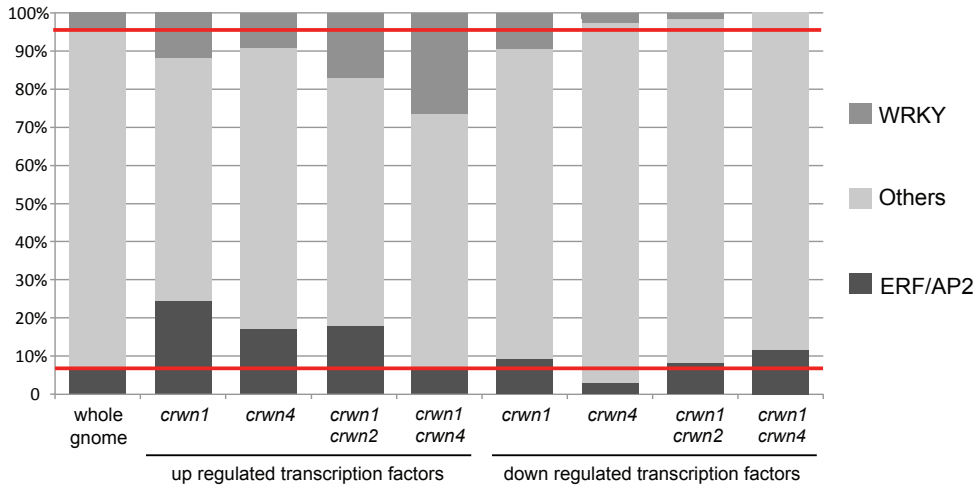


Figure 6B

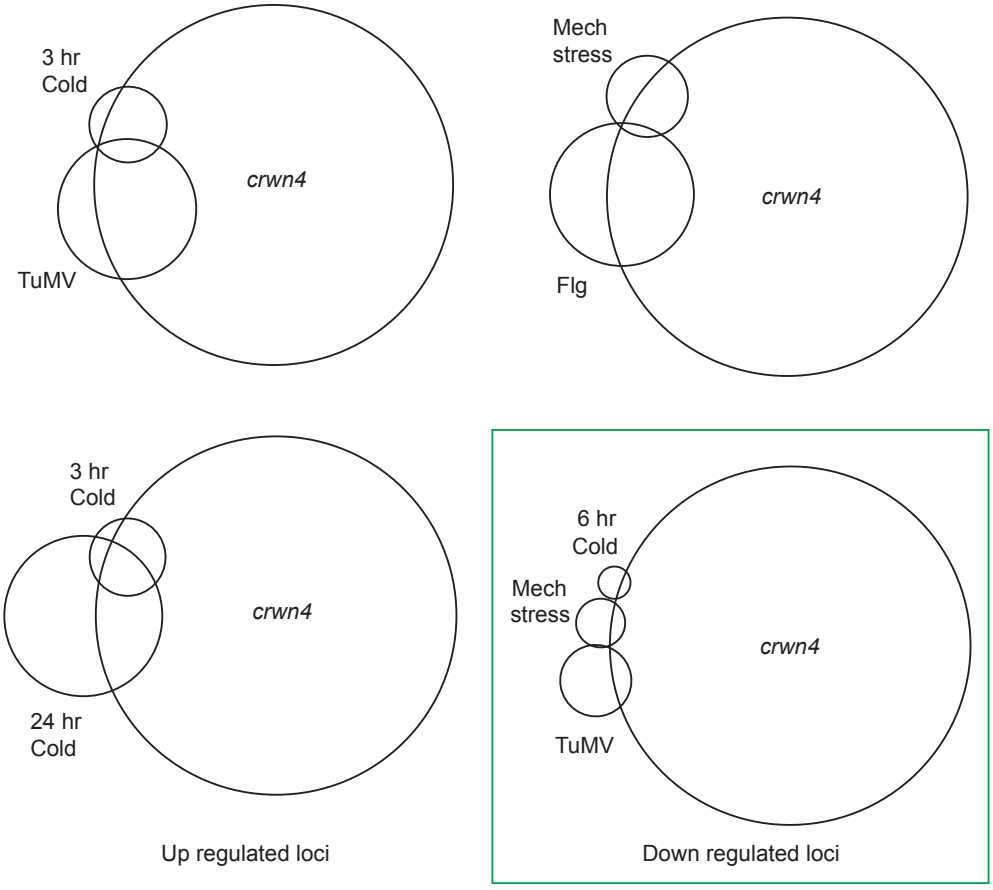


Figure 3.6 The activation of stress response pathways in *crwn* mutants

Figure 3.6 Panel A uses a percentage bar graph to exhibit the enrichment of ERF/AP2 and WRKY transcriptional factors among mis-expressed loci in *crwn1, crwn4, crwn1 crwn2*, and *crwn1 crwn4* mutants. A baseline percentage was calculated as follows: total number of ERF/AP2 family transcription factors / total number of transcription factors in the whole genome. This calculation was also applied to WRKY family. Panel B uses Venn diagrams to display the overlap of significantly mis-regulated loci among *crwn4* and various stressed wild type plants. Three comparisons among up-regulated loci were made; and a fourth one in the green box displays several multi sample comparisons of down-regulated loci between *crwn4* and several stressed wild type plants. There were no overlapping loci between the three-hour cold treatment experiment and the *crwn4* sample (not shown); a few loci overlapped among the six-hour cold treatment, the mechanical stress treatment and TuMV infection.

Supplementary Table 3.1A The activation of stress response pathways

	whole genome total	up regulation				down regulation			
		crwn1	*crwn4*	*crwn1 crwn2*	*crwn1 crwn4*	*crwn1*	*crwn4*	*crwn1 crwn2*	*crwn1 crwn4*
TXF total	1677	93	251	195	76	32	189	194	69
ERF/AP2	117	23	43	35	5	3	6	16	8
WRKY	75	11	23	33	20	3	5	3	0
other TXF	1485	59	185	127	51	26	178	175	61
ERF/AP2 // TXF total	6.98%	24.73%	17.13%	17.95%	6.58%	9.38%	3.17%	8.25%	11.59%
WRKY / TXF total	4.47%	11.83%	9.16%	16.92%	26.32%	9.38%	2.65%	1.55%	0.00%

Supplementary Table 3.1B The activation of stress response pathways

Supplemental Table 1B Stress response in *crwn* - ERF/AP2 family

ERF/AP2 loci	type	class	wild type	*crwn1 crwn2*	*crwn4*	*crwn1*
AT1G63030	DREB	A-1	0.14	3.94	3.78	3.57
AT4G25480	DREB	A-1	1.14	2.73	2.99	
AT4G25470	DREB	A-1	0.85	5.74	4.46	
AT4G25490	DREB	A-1	0.89	5.71	3.34	
AT5G51990	DREB	A-1	0.45	7.03	4.03	
AT1G12610	DREB	A-1	2.68	4.65	3.93	2.47
AT3G11020	DREB	A-2	3.09	0.85		
AT5G05410	DREB	A-2	3.84	3.80	1.95	0.97
AT2G38340	DREB	A-2	0.06			
AT2G40350	DREB	A-2	0.29			
AT5G18450	DREB	A-2	0.00			
AT2G40340	DREB	A-2	0.89	1.24		
AT1G75490	DREB	A-2	1.71			
AT3G57600	DREB	A-2	8.93	-0.97		
AT2G40220	DREB	A-3	0.00			
AT1G01250	DREB	A-4	9.69	-2.22	-2.22	-1.28
AT1G12630	DREB	A-4	0.12			
AT1G33760	DREB	A-4	0.78	7.24	4.19	1.83
At1G63040	DREB	A-4	2.08			
AT1G71450	DREB	A-4	0.00			
AT1G77200	DREB	A-4	0.14			
AT2G25820	DREB	A-4	5.12	-1.50		
AT2G35700	DREB	A-4	2.64	-2.78	-1.55	
AT2G36450	DREB	A-4	0.00			
AT2G44940	DREB	A-4	15.50	-1.72		
AT3G16280	DREB	A-4	1.13			
AT3G60490	DREB	A-4	3.42	-1.80		
AT4G16750	DREB	A-4	5.69			
AT4G32800	DREB	A-4	121.90		0.65	
AT5G11590	DREB	A-4	4.35	-2.48		
AT5G25810	DREB	A-4	16.86	-3.58	-2.80	-1.40
AT5G52020	DREB	A-4	0.45	4.18	3.93	2.61
AT1G19210	DREB	A-5	0.99	5.22	2.27	
AT1G21910	DREB	A-5	2.87	3.30	1.79	
AT1G22810	DREB	A-5	1.57		2.63	
AT1G44830	DREB	A-5	1.13			
AT1G46768	DREB	A-5	14.80	-0.83		
AT1G71520	DREB	A-5	0.09			
AT1G74930	DREB	A-5	6.49	5.14	3.34	1.72
AT1G77640	DREB	A-5	0.91	4.91	2.91	1.49
AT2G23340	DREB	A-5	51.15	-1.10		
AT3G50260	DREB	A-5	5.77	3.73	3.75	2.16
AT4G06746	DREB	A-5	16.37			
AT4G31060	DREB	A-5	3.94			
AT4G36900	DREB	A-5	16.72		2.70	2.12
AT5G21960	DREB	A-5	1.13	4.63	2.12	
AT5G67190	DREB	A-5	16.94		0.76	
AT4G28140	DREB	A-6	34.71		2.55	0.87
AT1G78080	DREB	A-6	223.67			
AT4G39780	DREB	A-6	12.91	0.97	0.84	
AT1G36060	DREB	A-6	0.24			
AT4G13620	DREB	A-6	0.02			
AT1G64380	DREB	A-6	51.12			
AT2G22200	DREB	A-6	15.26	-2.64	-0.81	
AT5G65130	DREB	A-6	0.39			

ERF/AP2 loci	type	class	wild type	*crwn1 crwn2*	*crwn4*	*crwn1*
AT1G03800	ERF	B-1	0.06			
AT1G12980	ERF	B-1	0.00		inf	
AT1G24590	ERF	B-1	0.00			
AT1G28160	ERF	B-1	0.00			
AT1G28360	ERF	B-1	0.14			
AT1G28370	ERF	B-1	4.04	6.34	4.77	2.40
AT1G50640	ERF	B-1	52.37	0.73	0.87	
AT1G53170	ERF	B-1	17.09	2.15	1.98	0.86
AT1G80580	ERF	B-1	0.00			
AT3G15210	ERF	B-1	90.93	2.40	1.41	
AT3G20310	ERF	B-1	22.76	1.57	1.34	
AT5G13910	ERF	B-1	0.04			
AT5G18560	ERF	B-1	0.00			
AT5G44210	ERF	B-1	3.68	-1.19	0.98	
AT1G72360	ERF	B-2	3.26		1.35	1.26
AT1G53910	ERF	B-2	108.09			
AT2G47520	ERF	B-2	0.04			
AT3G14230	ERF	B-2	74.20			
AT3G16770	ERF	B-2	158.64		1.21	0.96
AT1G04370	ERF	B-3	0.00			
AT1G06160	ERF	B-3	1.71	2.99	3.35	2.57
AT2G31230	ERF	B-3	4.79			
AT2G44840	ERF	B-3	4.47	3.57	3.10	1.79
AT3G23220	ERF	B-3	0.00			
AT3G23230	ERF	B-3	0.54	2.72	2.99	
AT3G23240	ERF	B-3	0.48	4.17	2.60	
AT4G17490	ERF	B-3	1.40	4.79	2.28	1.42
AT4G17500	ERF	B-3	9.75	3.28	2.84	2.62
AT4G18450	ERF	B-3	0.00			
AT4G34410	ERF	B-3	7.83	4.91	4.96	3.66
AT5G07580	ERF	B-3	59.75			
AT5G43410	ERF	B-3	0.00			
AT5G47220	ERF	B-3	15.54	3.15	3.92	3.59
AT5G47230	ERF	B-3	3.23	3.02	2.14	1.36
AT5G51190	ERF	B-3	2.93	3.18	1.55	
AT5G61590	ERF	B-3	73.06			0.97
AT5G61600	ERF	B-3	9.82	2.86		
AT5G07310	ERF	B-4	1.22	-1.90	-1.55	
AT2G33710	ERF	B-4	0.28	2.22	1.74	
AT5G61890	ERF	B-4	16.67	-2.66		-0.84
AT5G64750	ERF	B-4	44.74		1.37	
AT5G50080	ERF	B-4	0.00			
AT1G43160	ERF	B-4	141.77	-2.77	-1.58	-1.34
AT5G13330	ERF	B-4	6.82			
AT4G27950	ERF	B-5	3.22	-1.36		
AT5G53290	ERF	B-5	1.28			
AT4G23750	ERF	B-5	3.26			
AT1G22985	ERF	B-5	4.74			
AT1G71130	ERF	B-5	18.19			
AT4G11140	ERF	B-5	0.81			
AT1G15360	ERF	B-6	0.00			
AT1G25470	ERF	B-6	6.72			
AT1G49120	ERF	B-6	0.00			
AT1G68550	ERF	B-6	27.91			
AT2G20350	ERF	B-6	0.26			
AT3G25890	ERF	B-6	22.64		1.28	0.98
AT5G11190	ERF	B-6	0.17			
AT5G19790	ERF	B-6	0.30			
AT5G25190	ERF	B-6	6.04			
AT5G25390	ERF	B-6	0.13			
AT5G67000	ERF	B-6	0.07			
AT5G67010	ERF	B-6	0.16			

Supplementary Table 3.1C The activation of stress response pathways

Supplemental Table 1C Stress response in *crwn* - WRKY family

Locus ID	gene name	*crwn1*	*crwn4*	*crwn1 crwn2*	*crwn1 crwn4*	*crwn1*	*crwn4*	*crwn1 crwn2*	*crwn1 crwn4*
AT2G04880	WRKY01								
AT5G56270	WRKY02			WRKY02					
AT2G03340	WRKY03						WRKY03		
AT1G13960	WRKY04								
AT5G46310	WRKY05								
AT1G62300	WRKY06		WRKY06	WRKY06	WRKY06				
AT4G24240	WRKY07			WRKY07					
AT5G46350	WRKY08				WRKY08				
AT1G68150	WRKY09								
AT1G55600	WRKY10								
AT4G31550	WRKY11		WRKY11	WRKY11					
AT2G44745	WRKY12								
AT4G39410	WRKY13								
AT1G30650	WRKY14								
AT2G23320	WRKY15	WRKY15	WRKY15	WRKY15	WRKY15				
AT5G45050	WRKY16								
AT2G24570	WRKY17								
AT4G31800	WRKY18	WRKY18	WRKY18	WRKY18	WRKY18				
AT4G12020	WRKY19								
AT4G26640	WRKY20								
AT2G30590	WRKY21			WRKY21					
AT4G01250	WRKY22	WRKY22	WRKY22	WRKY22					
AT2G47260	WRKY23								
AT5G41570	WRKY24								
AT2G30250	WRKY25	WRKY25	WRKY25	WRKY25	WRKY25				
AT5G07100	WRKY26	WRKY26	WRKY26	WRKY26	WRKY26				
AT5G52830	WRKY27					WRKY27	WRKY27	WRKY27	
AT4G18170	WRKY28		WRKY28					WRKY28	
AT4G23550	WRKY29		WRKY29						
AT5G24110	WRKY30	WRKY30	WRKY30	WRKY30					
AT4G22070	WRKY31								
AT4G30935	WRKY32						WRKY32	WRKY32	
AT2G38470	WRKY33	WRKY33	WRKY33	WRKY33	WRKY33				
AT4G26440	WRKY34								
AT2G34830	WRKY35								
AT1G69810	WRKY36			WRKY36	WRKY36				
AT5G22570	WRKY38			WRKY38	WRKY38				
AT3G04670	WRKY39								
AT1G80840	WRKY40	WRKY40	WRKY40	WRKY40	WRKY40				
AT3G32090	WRKY40								
AT4G11070	WRKY41			WRKY41					

Supplementary Table 3.1D The activation of stress response pathways

Supplemental Table 1D Stress response in *crwn* - pathway overlap

wild type	*crwn1*	*crwn4*	*crwn1 crwn2*	*crwn1*	*crwn4*	*crwn1 crwn2*	TAIR 10	*crwn1*	*crwn4*	*crwn1 crwn2*	*crwn1*	*crwn4*	*crwn1 crwn2*
27416	781	2922	2459	534	2632	2279	*crwn* mutants	2.85%	10.66%	8.97%	1.95%	9.60%	8.31%
161	47	96	93	4	26	19	mechanical	29.19%	59.63%	57.76%	2.48%	16.15%	11.80%
60	25	42	47	1	5	3	mec top 60	41.67%	70.00%	78.33%	1.67%	8.33%	5.00%
127	61	107	102	2	3	3	3 h cold	48.03%	84.25%	80.31%	1.57%	2.36%	2.36%
194	74	134	124	4	6	22	6 h cold	38.14%	69.07%	63.92%	2.06%	3.09%	11.34%
580	75	211	159	21	66	86	24 h cold	12.93%	36.38%	27.41%	3.62%	11.38%	14.83%
536	127	328	405	10	33	35	virus	23.69%	61.19%	75.56%	1.87%	6.16%	6.53%
326	42	117	137	9	63	44	drought	12.88%	35.89%	42.02%	2.76%	19.33%	13.50%
1193	91	328	350	20	61	48	virus + drought	7.63%	27.49%	29.34%	1.68%	5.11%	4.02%
2859	420	974	919	88	221	201	heat	14.69%	34.07%	32.14%	3.08%	7.73%	7.03%
3109	166	533	481	89	343	285	virus + drought + heat	5.34%	17.14%	15.47%	2.86%	11.03%	9.17%
785	106	296	257	60	152	142	DC3000-6	13.50%	37.71%	32.74%	7.64%	19.36%	18.09%
600	92	251	192	36	98	97	DC3000-8	15.33%	41.83%	32.00%	6.00%	16.33%	16.17%
489	97	291	282	18	38	34	E coli - flg	19.84%	59.51%	57.67%	3.68%	7.77%	6.95%

wild type	*crwn1*	*crwn4*	*crwn1c rwn2*	*crwn1*	*crwn4*	*crwn1 crwn2*	TAIR 10	*crwn1*	*crwn4*	*crwn1 crwn2*	*crwn1*	*crwn4*	*crwn1 crwn2*
27416	534	2632	2279	781	2922	2459	*crwn* mutants	2.85%	10.66%	8.97%	1.95%	9.60%	8.31%
44	5	10	7	1	10	4	mechanical	2.27%	22.73%	9.09%	11.36%	22.73%	15.91%
4	0	0	0	0	0	0	3 h cold	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
21	0	4	2	0	0	0	6 h cold	0.00%	0.00%	0.00%	0.00%	19.05%	9.52%
265	21	89	68	7	19	24	24 h cold	2.64%	7.17%	9.06%	7.92%	33.58%	25.66%
125	4	21	16	2	9	13	virus	1.60%	7.20%	10.40%	3.20%	16.80%	12.80%
118	6	19	31	13	26	11	drought	11.02%	22.03%	9.32%	5.08%	16.10%	26.27%
482	10	116	100	16	71	55	virus + drought	3.32%	14.73%	11.41%	2.07%	24.07%	20.75%
1623	88	393	341	52	173	124	heat	3.20%	10.66%	7.64%	5.42%	24.21%	21.01%
3178	149	733	637	161	472	403	virus + drought + heat	5.07%	14.85%	12.68%	4.69%	23.06%	20.04%
893	64	308	274	50	135	117	DC3000-6	5.60%	15.12%	13.10%	7.17%	34.49%	30.68%
1056	78	329	256	79	203	196	DC3000-8	7.48%	19.22%	18.56%	7.39%	31.16%	24.24%
621	59	263	213	19	55	49	E coli - flg	3.06%	8.86%	7.89%	9.50%	42.35%	34.30%

Supplementary Table 3.1 Supplementary Table 3.1A displays the statistics of mis-expressed ERF/AP2 and WRKY transcription factors in *crwn* mutants. The locus ID and RPKM value are listed for these enriched loci in Supplementary Table 3.1B (ERF/AP2) and Supplementary Table 3.1C (WRKY). Supplementary Table 3.1D shows the enrichment of *crwn* mis-regulated loci in various experimentally stressed wild type plants.

To investigate the activation of stress response pathways further, I compared the transcriptomic data between *crwn* mutants and wild type plants that were experimentally subjected to either abiotic or biotic stresses, including cold stress, mechanical stress, virus infection, bacteria treatment, heat stress and drought treatment [147-150] (Supplementary Table 3.1). The Venn diagrams in Figure 3.6B summarize the common transcriptional targets among different treatments and the *crwn4* mutant. A very significant percentage (>75%) of up-regulated targets in the response to either three hours of cold treatment or TuMV (Turnip mosaic virus) infection (10 minutes post infection) belong to the up-regulated loci in *crwn4* mutants, consistent with the previously noted up-regulation of ERF/AP2 and WRKY transcription factor families under these conditions. A slightly lower percentage (~60%) of up-regulated loci in the experiment involving brief mechanical stress were shared with the *crwn4* mutant (Supplementary Table 3.1D; an 80% overlap of the top 60 extremely over expressed loci). However, a lower percentage of overlap was seen between crwn4 mutants and stressed wild type plants among the down-regulated loci (bottom panel of Figure 3.6, and Supplementary Table 3.1D). These data indicate that *crwn* mutants preferentially activate some stress response pathways.

Related to defense responses, several members of the wall associated protein kinase family [151] were up-regulated, and cell wall and extra-cellular proteins were enriched among down-regulated loci, including receptor proteins sensing various environment stimuli and enzymatic proteins regulating cell wall stiffness and expansion. These changes were most notable in *crwn1* and *crwn2* single mutants,

indicating that genes ecnoding cell wall and extra-cellular proteins are transcriptionally sensitive to even mild disturbances in the nuclei due to the loss of CRWN proteins.

Nuclear proteins mis-expressed in *crwn* mutants

I also attempted to understand how the transcriptional mis-regulation of loci might relate to the altered nuclear morphology seen in *crwn* mutants. I focused my examination on nuclear-localized proteins in three categories. First, I considered proteins in the nuclear envelope, nuclear pore complex, and putative nuclear matrix. In this category, one notable locus is the nucleoporin auto-peptidase encoded by *At1g59660*, which is up-regulated by five fold and three fold in *crwn1 crwn2* and *crwn1 crwn4* mutants, respectively. This up-regulation is unusual as transcripts for other nucleoporins were unaffected, with the exception of *At2g05120* (Nup133/Nup155-like), which is mildly down-regulated (less than two fold) in the *crwn4* mutant. Transcripts encoding two other nuclear membrane components, *At3g13360* (WIP3) and *At3g63000* (NPL41, Nuclear Pore Localization Protein 4-like protein 1) were also slightly down-regulated (less than two fold) in the *crwn4* mutant.

The second class of nuclear proteins I examined are ones important in chromosome organization, including SMC (Structural Maintenance of Chromosomes) complex components and proteins involved in chromatin condensation and sister chromatid cohesion [100], as well as nuclear-localized actin-related proteins (AtARP), and

putative chromosome-interacting AT hook family proteins. Among SMC members *crwn4* and *crwn1 crwn2* mutants, while transcription for the *SMC1* and *SMC3* were not disturbed. The SMC5/6 complexes, which promote sister chromatid cohesion for DNA repair through homologous recombination in somatic cells [102], were severely disrupted. The SMC6A subunit was significantly under-expressed in *crwn1, crwn4,* and *crwn1 crwn2* mutants, as were non-SMC proteins necessary for sister chromatin cohesion (SYN3, ETG1 and CTF18, Nse1) particularly in *crwn4*. These changes, together with the alterations in condensin and cohesin formation, might cause the chromocenter diffusion seen in *crwn4* nuclei ([140], Chapter2, Table 3.4A). On the other hand, SMC6B and non-SMC members in the complex Nse4 and HEB2, were up-regulated in *crwn1 crwn2* and *crwn1 crwn4* mutants (Table 3.4B), which could be related to the chromocenter aggregation seen in these mutants. Among the actin-related proteins, nuclear localized AtARP8 (encoded by *At5g56180*) was up-regulated in *crwn4*. Among the approximately 50 members in the AT hook protein family, four members were down-regulated in the *crwn4* mutant, including *At1g63470, At1g63480, At2g45850*, and *At3g04590* (Table 3.4A) .

Finally, I considered proteins serving basic nuclear functions in DNA and RNA metabolism. In total, there are 29 genes up-regulated and 54 genes down-regulated genes in this class in *crwn1 crwn2* mutants. In *crwn4* mutants, there are 34 genes up-regulated and 120 genes down-regulated belonging to this class. Noticeably, the loss of CRWN1 protein alone leads to many changes in this category, although small and round nuclei are the only obvious phenotypes observed in *crwn1* mutant. First, a set of

endo-reduplication regulators was significantly mis-expressed, including LGO and TCP15, which were down-regulated, and KRP1, which was up-regulated. These alterations should prohibit the cells from entering the endo-cycles [152]. Meanwhile, the negative regulator of endo reduplication, cycle2,3A is significantly down-regulated, possibly as a compensation for mis-expression of LGO, TCP15, and KRP1. (Table 3.4C) The more severe *crwn1 crwn2* and *crwn4* mutants also exhibited a significantly reduced expression level of SIM and KRP2, the key positive regulators of endo cycles (Table 3.4C) [153]. Second, several key epigenetic modifiers (MET1, VIM1, SUVR2) were expressed at a lower level in *crwn1*, and this down-regulation was significantly expanded to many other proteins in severe *crwn* mutants, including a 50% reduction of *DDM1* expression in *crwn4* mutants (Table 3.4E). The mis-expression of the epigenetic modifiers for DNA methylation was coupled with transcriptional changes of histone modification regulators. These changes are correlated with visible changes in epigenetic mis-regulation in *crwn* mutants (Figure 3.2 and Table 3.3). A few examples of nucleosome assembly components (histone H2B, H2A) were under-expressed in *crwn1*, and roughly 10 to 20 of histone subunits of H2A/B, H3, H4 were added into this down-regulated list in *crwn1 crwn2* and *crwn4* mutants (Table 3.4G and Table 3.4H). These results suggest that *crwn* mutations might cause open chromatin configurations. In addition, DNA polymerase and many other subunits of the DNA replication machinery (DNA pol alpha, MCM2, MCM5, ANP3, SMC6A, TPX2, RNR1, RPA70D) are expressed at a lower level in *crwn1* mutants, and this list grew in *crwn1 crwn2* (*e.g.*, cyclinB2, involved in mitotic cell cycles) and *crwn4* mutants (ATX1, ORC complexes, DNA pol A, I, III, alpha, gamma, Y family,

etc.), consistent with hindered endo-cycles, decreased endopolyploidy and reduced nuclear area. These changes might also affect regular mitotic cycles (Table 3.4D). Lastly, an over-expression of DNA repair enzymes (DNA glycosylase, SNM1, BARD1, BBX2) participating in base excision, non-homologous and homologous recombination mediated DNA repair was observed in *crwn1* mutants. In one particularly striking example, a DNA glycosylase encoded by *At1g75230* was up-regulated 10 fold in *crwn4*, indicating an activation of base excision DNA repair. These changes were associated with a broad activation of other DNA repair factors (DAN1, RAD21, RAD23B, PARP2, CEN2, REV1, RAD51, XRI1, RECQ2) in more severe *crwn* mutants (Table 3.4F). Taken together, these changes suggest that the loss of CRWN proteins interrupt various basic nuclear processes, including DNA replication, endo-cycle regulation, epigenetic modification, nucleosome assembly and DNA repair.

Table 4A Nuclear envelope proteins, AT hook containing proteins, Actin related proteins

Locus ID	gene name	categories	wild type	*crwn1*	*crwn4*	*crwn1 crwn2*	*crwn1 crwn4*	*crwn2*
AT1G59660	Nucleoporin autopeptidase	Nuclear envelope	2.09			11.57	6.97	
AT1G79280	NUA		19.23					12.38
AT2G05120	Nup133/Nup155-like nucleoporin		3.61		2.38			
AT3G13360	WIP3		33.38		21.43			
AT3G63000	NPL41		32.75		23.31			
AT1G63470		AT-hook	5.59		2.46			
AT1G63480			5.63		2.14			
AT2G45850			17.72		10.60			
AT3G04590			27.88		19.50			
AT1G13180		At ARP	7.85		5.24			
AT5G56180			9.96		16.40			

Table 4B SMC complexes subunit and interacting proteins

Locus ID	gene name	complex	wild type	*crwn1*	*crwn4*	*crwn1 crwn2*	*crwn1 crwn4*	*crwn2*
AT3G47460	SMC2	codensin	0.53	0.45	0.54	0.40	0.53	0.48
AT5G61460	SMC6B	SMC5/6	2.03	2.25	2.05	3.15	2.91	1.93
AT3G16730	HEB2	condensin	0.93	0.91	0.94	1.77	1.28	1.04
AT1G51130	Nse4	SMC5/6	1.22	1.04	1.17	1.71	1.52	1.39
AT5G15920	SMC5	SMC5/6	3.60	3.60	3.44	3.58	4.17	3.01
AT2G27170	SMC3	cohesin	13.13	12.13	11.54	11.99	16.37	10.13
AT3G54670	SMC1	cohesin	7.68	7.67	6.62	7.45	10.70	*6.32*
AT3G57060		condensin	1.12	1.07	0.85	0.81	1.05	0.96
AT5G21140	Nse1	SMC5/6	1.70	1.61	1.16	1.40	1.75	*1.53*
AT5G62410	SMC2	codensin	1.01	0.64	0.54	0.63	1.03	0.72
AT3G23890	TOPII	cell division	0.55	0.23	0.25	0.20	0.33	0.30
AT5G48600	SMC4	codensin	1.14	0.58	0.50	0.67	0.92	0.68
AT1G04730	CTF18	cohesion	0.57	0.32	0.24	0.48	0.66	0.53
AT2G40550	ETG1	cohesion	1.84	0.79	0.74	2.03	1.58	1.39
AT3G49250	DMS3/IDN1	DDR	3.57	2.23	1.40	1.90	2.80	*3.38*
AT5G07660	SMC6A	SMC5/6	1.11	0.27	0.15	0.34	0.88	0.62
AT3G59550	SYN3	chromatid	0.43	0.09	0.05	0.10	0.26	0.23

Table 4C Endo-reduplication regulators

locus ID	gene name	widlt type	*crwn4*	*crwn1 crwn2*	*crwn1 crwn4*	*crwn1*
AT3G59550	RAD21, SYN3	0.44	0.05	0.10		
AT1G15570	cyclin A2;3	1.58	0.34	0.28		0.48
AT2G42260	PYM, UVI4	0.48	0.11			
AT3G10525	LGO	131.39	31.52	42.41		60.51
AT5G67100	ICU2, DPA	1.08	0.34			
AT3G60840	MAP65	0.71		0.24		
AT1G03780	TPX2	0.42		0.15		
AT3G48160	DEL1	0.56	0.21			
AT3G50630	KRP2	10.68	4.07	5.24		
AT1G69690	TCP15	34.54	13.38	9.43		21.07
AT1G33240	GTL1	39.15	15.50	24.22		
AT2G40550	ETG1	1.87	0.75			
AT5G04470	SIM	18.15	9.24	6.46	9.41	
AT2G32710	KRP4	12.40	7.34			
AT1G20330	CVP1	99.75	61.97	59.85		
AT2G23430	KRP1	38.05	24.35			
AT4G22910	CCS52A1	8.25	5.30			
AT1G77390	CYCA1	0.95	3.77	2.88		
AT3G19150	KRP6	17.74	56.15	52.34	27.92	37.91
AT1G75950	ASK1	266.31	378.61			
AT3G08690	UBC11	51.64	70.77	79.98		
AT5G10440	CYCD4	0.69		1.63		
AT2G21550		0.60		1.42		

Table 4D DNA replication machinery

locus ID	gene name	wild type	*crwn4*	*crwn1 crwn2*	*crwn1 crwn4*	*crwn1*
AT4G38680	CSP2				48.86	
AT5G37630	EMB2656				0.63	
AT5G43990	SUVR2	1.12	0.32	0.52	0.58	
AT5G45720	DNA Pol III	1.52	0.78	0.23	0.88	
AT5G07660	SMC6A	1.13	0.15			0.27
AT1G20720	RAD3-like	0.43	0.06			
AT3G02820		1.82	0.32			
AT5G61000	RPA70D	1.84	0.32	0.70		0.55
AT3G14890		4.87	1.06	1.18		
AT2G20980	MCM10	0.72	0.16	0.24		
AT3G06030	ANP3	2.27	0.52	0.28		0.89
AT5G41880	DNA Pol A	1.42	0.33			
AT1G75150		0.64	0.16			
AT5G44635	MCM6	1.00	0.25			
AT3G12170	perone DnaJ-don	0.76	0.19			
AT4G12620	ORC1B	0.50	0.13			
AT5G15510	PX2 protein famil	1.48	0.43	0.43		0.52
AT5G13060	ABAP1	0.49	0.15	0.19		
AT3G23740		0.68	0.22			
AT1G67630	DNA Pol alpha 2	1.13	0.36			
AT5G52950		0.54	0.18			
AT1G44900	MCM2	1.50	0.51			0.61
AT2G24970		1.75	0.60			
AT2G07690	MCM5	1.81	0.63			0.75

Table 4D continue

AT2G37560	ORC2	0.49	0.18			
AT2G16440	MCM4	1.22	0.46			
AT3G49250	DMS3	3.62	1.42			
AT4G14770	TCX2	0.95	0.39			
AT1G04730	CTF18	0.58	0.24			
AT5G16690	ORC3	0.66	0.28			
AT5G48600	SMC3	1.16	0.50			
AT1G63470	AT hook	5.59	2.46			
AT3G23890	TOPII	0.55	0.26	0.21		
AT2G31650	ATX1	1.25	0.60			
AT5G23420	HMGB6	5.82	2.82			
AT2G21790	RNR1	9.54	4.70			4.68
AT3G42660		1.54	0.76			
AT4G18820	DNA Pol III	9.49	4.74	6.09		
AT5G13960	SUVH4	3.50	1.79			
AT2G36200		0.67	0.35	0.23		
AT5G62410	SMC4	1.03	0.54			
AT4G24790	DNA Pol III	1.88	1.00			
AT4G36180		2.46	1.39			
AT3G54750		3.06	1.75			
AT2G16390	DRD1	2.45	1.45			
AT1G14460	DNA pol III	3.86	2.41			
AT3G20540	DNA Pol I	10.32	6.64			
AT3G22780	TSO1	9.96	6.61			
AT1G50840	DNA Pol gama2	16.43	11.68			
AT1G76310	CYCLIN B2	0.99		0.27		
AT3G20150		0.44		0.12		
AT4G33400		2.71		1.34		
AT2G06510	RPA1A				6.25	
AT3G27060	TSO2	19.89		48.25	36.61	
AT4G02070	MSH6				2.59	
AT4G19130	RPA1 related				0.88	
AT4G37490	CYCB1				1.56	
AT1G33420		2.55	4.52	7.14		
AT1G80190	PSF1	4.03	9.56			
AT2G21660	GRP7	1729.44	3321.58			
AT4G36020	CSDP1	16.98	28.78	29.99		
AT5G09790	ATXR5/SDG15	2.40		4.94		
AT5G49570	PNG1	10.94	16.21			

Table 4E DNA methylation modifiers

locus ID	gene name	wild type	*crwn4*	*crwn1 crwn2*	*crwn1 crwn4*	*cwn1*
AT5G43990	SUVR2	1.12	0.32	0.52	0.58	0.42
AT4G36180	LLR kinase family	2.46	1.39		1.56	
AT5G25590		3.27	0.72	0.71	1.21	1.17
AT4G37750	CKC1	1.31	0.55		0.51	
AT2G28290	CHR3	12.71	8.79			
AT3G12710	DNA glycosylase				1.61	

Table 4E continue

AT2G36490	ROS1	16.17	11.83	8.15		
AT1G04020	BARD1	1.11	0.21	0.35		
AT5G08020	RPA70B	0.88	0.18			
AT5G57970	DNA glycosylase	2.40	0.90	0.96		
AT5G07660	SMC6A	1.13	0.15	0.35		0.27
AT3G06030	ANP3	2.27	0.52	0.28		0.89
AT5G15510	TPX2 protein family	1.48	0.43	0.43		0.52
AT3G23740		0.68	0.22			
AT1G67630	DNA Pol alpha 2	1.13	0.36			
AT5G52950		0.54	0.18			
AT1G44900	MCM2	1.50	0.51			0.61
AT2G07690	MCM5	1.81	0.63			0.75
AT2G16440	MCM4	1.22	0.46			
AT3G49250	DMS3	3.62	1.42	1.94		
AT5G48600	SMC3	1.16	0.50			
AT1G63470	AT hook motif	5.59	2.46			
AT2G31650	ATX1	1.25	0.60			
AT3G42660		1.54	0.76			
AT5G13960	SUVH4	3.50	1.79			
AT2G36200		0.67	0.35	0.23		
AT5G62410	SMC4	1.03	0.54			
AT3G54750		3.06	1.75			
AT2G16390	CHR35, DRD1	2.45	1.45			
AT3G22780	TSO1	9.96	6.61			
AT2G40550	ETG1	1.87	0.75			
AT3G14980	ROS4, IDM1	7.80	3.85			
AT4G29360		1.70	0.93	0.51		
AT3G12550	FDM3	5.73	3.16			
AT1G57820	ORTH2, VIM1	1.82	0.49	0.62		0.74
AT5G49160	MET1	2.18	0.76			0.98
AT2G23380	CLF	4.14	2.23			
AT5G04290	KTF1	6.45	3.65			
AT4G19020	CMT2	5.61	3.76			
AT3G23890	TOPII	0.56		0.21		
AT1G76310	CYCLIN B2	0.99		0.27		
AT3G20150		0.44		0.12		
AT4G33400		2.71		1.34		
AT1G03780	TPX2	0.42		0.15		
AT4G14200		1.47		0.35		
AT5G48360		2.41		1.25		
AT5G66750	DDM1	0.88	0.36			
AT3G57300	INO80				30.66	
AT4G37490	CYCB1				1.56	
AT5G04290	SPT5L				9.30	
AT5G04560	DME				15.88	
AT5G20850	RAD51				1.32	
AT1G15910	FDM1, IDP1	3.50	5.33	5.99		
AT1G63020	NRPD1	0.80		1.48		
AT1G80420	XRCC1	10.32	18.45	18.57		
AT2G30280	DMS4, RDDM4	6.81	10.76			
AT3G10010	DML2	2.86		5.60		
AT5G09790	ATXR5	2.39	4.97			
AT5G20320	DCL4	4.70	7.21	7.60		

Table 4F DNA repair pathways

locus ID	gene name	wild type	*crwn4*	*crwn1 crwn2*	*crwn1 crwn4*	*crwn1*
AT1G04020	BARD1	1.11	0.21	0.35		0.38
AT5G08020	RPA70B	0.88	0.18			
AT4G17760	damaged DNA binding	0.75	0.17			
AT5G24280	GMI1	1.82	0.64			
AT5G57970	DNA glycosylase	2.40	0.90	0.96		
AT5G64630	FAS2	1.46	0.55			
AT5G44680	DNA glycosylase	61.65	24.78	21.35		
AT3G46940	DUT1	4.19	1.71			
AT4G00020	BRCA2A	0.88	0.39	0.34		
AT2G01750	MAP70-3	3.70	2.33			
AT1G31360	RECQ2	6.06	4.02			
AT3G23580	RNR2A	28.67	19.23			
AT5G66050	wounding responsive	29.63	21.42			
AT2G36490	ROS1	16.17	11.83	8.15		
AT1G57820	ORTH2, VIM1	1.84		0.62		
AT3G49250	DMS3, IDN1	3.65		1.94		
AT5G07660	SMC6A	1.14	0.15	0.35		0.27
AT1G60930	RECQ4B	1.86		0.90		
AT1G02730	CSLD6, SOS6	1.77		0.52		
AT5G43080	Cyclin A3	1.55		0.48		
AT1G11190	BFN1, ENDO1	2.41		0.57	1.00	
AT1G02730	SOS6				0.88	
AT2G06510	RPA1A	3.97			6.25	
AT4G19130	RPA1 related	0.37			0.88	
AT4G21070	BRCA1	0.54			1.23	
AT4G37490	CYCB1	0.27			1.56	
AT5G03780	TRFL10	5.38			10.84	
AT5G04560	DME	11.03			15.88	
AT5G20850	RAD51	0.52			1.32	
AT5G24280	GMI 1	1.79			3.45	
AT5G49570	PNG1	10.78			16.00	
AT3G27060	TSO2	19.89		48.25	36.61	
AT4G02390	PARP2	0.98		4.60	7.79	
AT4G37010	CEN2	2.53		11.56	8.71	
AT5G44750	REV1	11.16		20.57	15.75	
AT5G48720	XRI	5.32		10.54	14.07	
AT1G78870	UBC13A	107.642	149.71			
AT5G58720		26.2747	73.254			
AT1G75230	DNA glycosylase	11.16	108.16	30.06		69.76
AT3G26680	SNM1	2.45	9.31	4.83		5.12
AT5G15850	BBX2	12.23	36.94			24.38
AT5G58720		26.27	73.25	68.49		
AT3G04620	DAN1	2.50	4.99			
AT1G80420	XRCC1	10.24	18.45	18.57		
AT1G79650	RAD23B	17.57	28.05			
AT1G78870	UBC13A	107.64	149.71			
AT3G10010	DML2	2.86		5.60		
AT3G12040		6.10		14.42		
AT5G20320	DCL4	4.70		7.60		

Table 3.4 This series of tables summarize the categories of nuclear proteins significantly mis-regulated in *crwn* mutants, including: nuclear envelope, AT hook and ARP (actin related protein) (Table 3.4A); SMC complexes relevant (Table 3.4B); endo-reduplication cycle regulators (Table 3.4C); DNA replication machinery (Table 3.4D); DNA methylation modifiers (Table 3.4E); DNA repair (Table 3.4F); histones (Table 3.4G), and histone modifiers (Table 3.4H). RPKM values were displayed for each locus in different genotypes. Red boxes contain up-regulated loci and green boxes contain down-regulated loci.

Transcriptional variation among crwn1 crwn2 replicates

To investigate the quality of these transcriptome datasets, I also checked the mRNA expression of CRWN genes, and evaluated the within-genotype variation through comparisons of each mutant sample to the three wild type samples individually. I found that in one of the three *crwn1 crwn2* replicates, both CRWN3 and CRWN4 mRNA expression level dropped significantly. Interestingly, examination of the distribution of sequence reads (Figure 3.7A) showed a dramatic reduction of CRWN3 and CRWN4 mRNA in the 5' region of the gene to the beginning of the sixth exon. This apparent post-transcriptional silencing of CRWN3 and CRWN4 is associated with variation of the *crwn1 crwn2* transcriptome (Figure 3.7B): the mis-regulated profile of the particular *crwn1 crwn2* replicate with partially silenced of CRWN3 and CRWN4 (12-1187), differs from the profiles of the other two *crwn1 crwn2* replicates (12-1221, and 12-1222). Despite this variability within the *crwn1 crwn2* samples, the relationships among the *crwn1 crwn2* genotype and other mutants were not altered. As represented in Figure 3.7B, all three replicates in *crwn1 crwn2* had a similar number of mis-expressed loci. Also, the average profile of *crwn1 crwn2* displayed a coherent group of targets among replicates (Supplementary Figure 3.1A). Moreover, the Venn diagram pattern of significantly mis-regulated loci between representative *crwn1 crwn2* replicates (12-1187 or 12-1222) and other *crwn* mutants (*crwn1*, *crwn2*, and *crwn4*) remained similar compared to the previous Venn diagram comparisons using the average *crwn1 crwn2* profile (Supplementary Figure 3.1B, C, D, and Figure 3.3). Therefore, the transcriptional variation induced by the silencing of *CRWN3* and

CRWN4 loci in one of the *crwn1 crwn2* replicates does not change the overall relationship between *crwn1 crwn2* and other *crwn* mutants on the transcriptomic level.

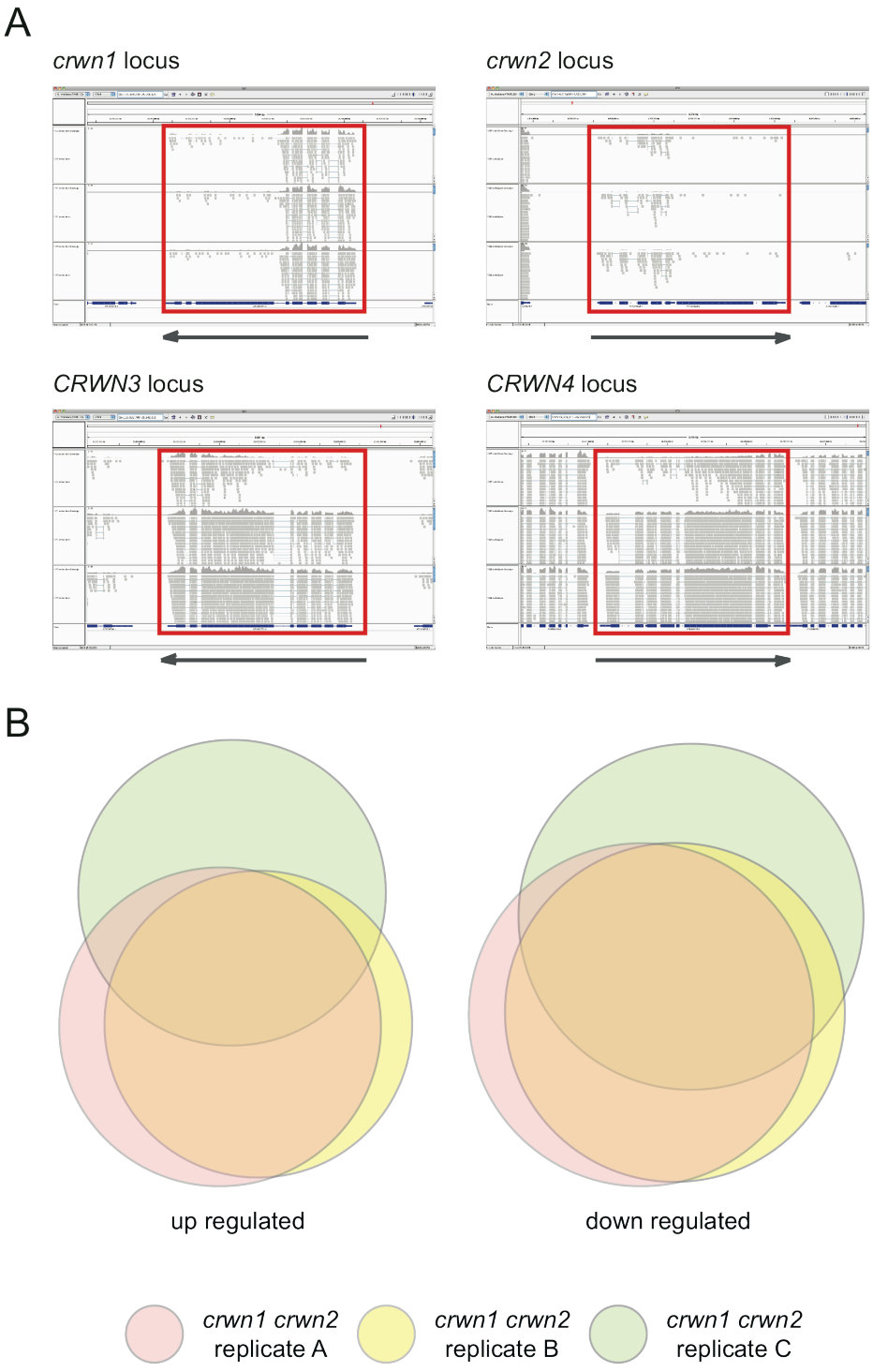
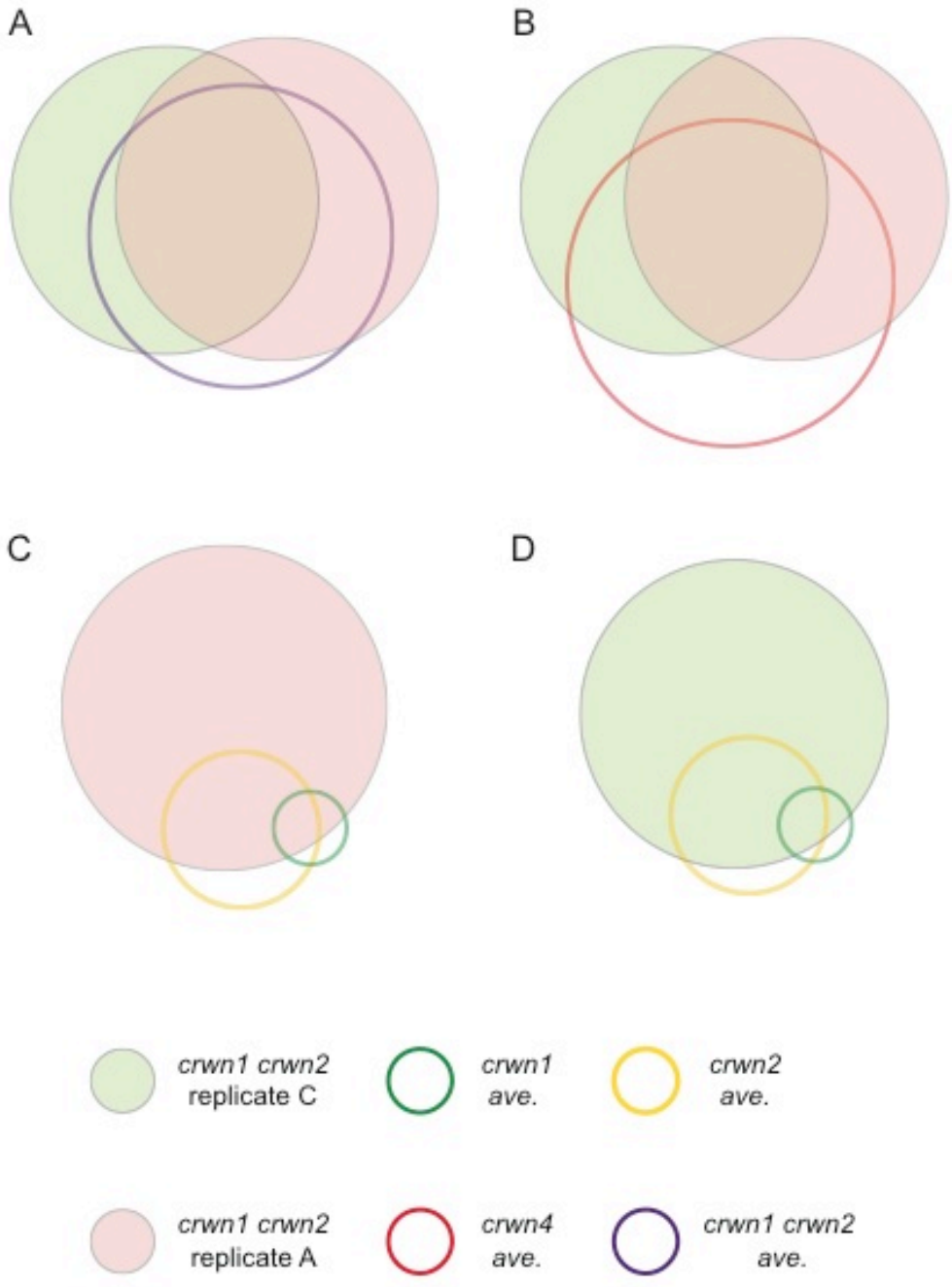


Figure 3.7 Partial silencing of CRWN3 and CRWN4 in one of the *crwn1 crwn2* replicates

Figure 3.7 Panel A displays the alignments of reads to the four *CRWN* loci in the three *crwn1 crwn2* replicates. In the *CRWN1* and *CRWN2* loci, the T-DNA insertions are positioned at the beginning of the sixth and longest exon. As expected, very few reads are present downstream of the insertion point, while the mRNA expression upstream of the T-DNA insertion is present. In two replicates, reads align with comparable depth across all exons of *CRWN3* and *CRWN4*. However, in replicate C, the mRNA upstream of the T- DNA insertion is largely reduced, and this reduction extends to the largest exon. In Panel B, The Venn diagrams illustrate significantly mis-regulated loci among the three *crwn1 crwn2* replicates. Replicate A, B, and C of *crwn1 crwn2* is represented in red, yellow and green, respectively. The size of the circle reflects the number of the loci, and the overlap between two circles reflects the number of shared mis-regulated loci between two *crwn1 crwn2* replicates.



Supplementary Figure 3.1 Partial silencing of CRWN3 and CRWN4 proteins in one of the three *crwn1 crwn2* replicates

Supplementary Figure 3.1 The Venn diagrams illustrate the significantly mis-regulated loci among representative *crwn1 crwn2* replicates and the average profile of different *crwn* mutants. Each circle with a distinct color represents a *crwn* mutant genotype or a *crwn1 crwn2* replicate. The size of the circle reflects the number of the loci, and the overlap between two circles reflects the number of shared mis-regulated loci between two *crwn* samples.

Discussion

In this study, I performed mRNA-seq profiling of selected *crwn* mutants in *A. thaliana*, and revealed significantly altered transcriptomes due to the loss of CRWN proteins. The comparison of the transcriptomic data from different *crwn* mutants (Figure 3.3) confirmed the synergistic interaction between *crwn1* and *crwn2* mutations, and uncovered the functional suppression between *crwn1* and *crwn4* mutations - both suggested by our initial genetic analysis [140]. Further, the transcription profiling indicated that *CRWN1*-like and *CRWN4*-like genes affect the expression of similar sets of genomic loci. It is likely that CRWN1 and CRWN4 proteins form a functional complex in the cell (see Chapter 4), providing a mechanistic explanation for the overlapping sets of transcriptional targets in *crwn1 crwn2* and *crwn4* mutants. Our previous interpretation of the phenotypic suppression between *crwn1* and *crwn4* mutations rested on the model that these proteins had at least partially antagonistic functions. However, my transcriptomic results point to a more straightforward hypothesis that CRWN1 and CRWN4 are working in concert and that the counter-intuitive genetic suppression result is possibly due to compensation by CRWN2 and/or CRWN3, which are up-regulated in *crwn1 crwn4* mutants (see Chapter 4). These considerations demonstrate the importance of combining transcriptomic data with traditional genetics analysis to aid in the interpretation of genetic interactions.

To understand how CRWN proteins affect transcription, I first examined whether the previously described phenotypic changes (Table 3.1) were directly correlated with the

severity of global transcriptional mis-regulation. First, *crwn1 crwn2* mutants had the most dwarfed stature and the smallest nuclei among the five tested *crwn* mutants, consistent with the most severe transcriptional changes observed in this double mutant. On the other hand, *crwn2* mutants, which resemble wild type morphologically and display only mild and variable effects on nuclear size, exhibited very limited changes in gene expression. At these extremes on the phenotypic spectrum, these two *crwn* genotypes illustrated an overall consistency between morphological and transcriptional alterations. However, this simple correlation does not always hold true. *crwn4* mutants resemble wild type plants morphologically, but their nuclear size is decreased by 50%; while *crwn1 crwn4* plants are slightly dwarfed, with shorter internodes, more branches, and a more extreme reduction in nuclear area to 25% of the wild type value (see Chapter 2). These facts contradict the observation that *crwn4* mutants had more widespread transcriptional changes that were largely suppressed in *crwn1 crwn4* mutants. Moreover, our previous cytological investigation of *crwn* mutants revealed that chromocenters were diffuse in *crwn4* nuclei but aggregated in *crwn1 crwn2* nuclei, yet these two genotypes displayed highly similar transcriptomes. Thus, the dispersion and aggregation of heterochromatin themselves are not associated with transcriptional activation or silencing, respectively. I note, however, that our analysis was limited to poly adenylated transcripts and that I was unable to detect transcription from heterochromatic repeats. Nonetheless, the loci mis-regulated in *crwn* mutants are well dispersed in the genome (data not shown), suggesting that an underlying disruption of nuclear organization in general is more likely responsible for these genome-wide transcriptional alterations. Although the overall morphological

changes seen in particular *crwn* mutants roughly correlate with the level of transcriptional disruption observed, no single phenotypic alteration predicts the extent of gene mis-expression at the global level.

I then tried to associate mis-expressed loci with corresponding alterations of nuclear function in *crwn* mutants, and considered which were likely primary changes versus downstream effects. Our transcriptomic data fits the original idea that CRWN proteins function as structural components of the nuclei because many of the alterations I noted affected nuclear proteins. I hypothesize that the loss of CRWN proteins primarily alters nuclear organization. These spatial changes could directly affect gene expression due to the mis-positioning of either chromosomes or various transcriptional machineries. The disruption of a scaffolding system in the nuclei could also impair many other nuclear functions, and result in transcriptional mis-regulation indirectly.

One category of nuclear functional change I investigated was epigenetic alteration in *crwn* mutants in an attempt to examine potential mechanistic connections between altered heterochromatin organization and genome-wide transcriptional changes. Our data uncovered a mild suppression of transposon silencing concentrated in the heterochromatic regions in *crwn* mutants (Figure 3.2). In addition, protein-coding genes regulated by epigenetic pathways are enriched among the significantly mis-expressed loci (Table 3.2), suggesting an impairment of epigenetic regulation in *crwn* mutants. The majority of the enrichment involved loci that are down-regulated in *met1* and *ddc,* or up-regulated in *rdd* mutations, rather than the more typical examples of

coding genes activated by a loss of DNA methylation. The small RNA-regulated genes in the *MET1* pathway were mostly undisturbed, implying that the RNA-directed CpG DNA methylation is not visibly affected by *crwn* mutations. Consistent with these observations, I observed a down-regulation of many nuclear proteins important for maintaining a dynamic balance of DNA methylation (Table 3.3). The mis-regulation of various epigenetic modification pathways provides a straightforward explanation for the broad but subtle reduction in epigenetic silencing in both transposon activation and enrichment of epigenetically-targeted coding genes. Another aspect reflecting open chromatin is the down-regulation of histone modifiers and under-expression of many histone proteins in *crwn* mutants, especially *crwn4*. These lowered activities may also attenuate the self-reinforced feedback loop for the dynamic maintenance of epigenetic modification. All of these changes could lead to downstream transcriptional mis-regulation. In the future, genome-wide profiling of different epigenetic marks will be needed to determine which genomic regions are primary targets for epigenetic regulation sensitive to a loss of CRWN protein function.

Overall, the mis-expression of epigenetic regulators is most significant in *crwn4* mutants, consistent with the dispersion of heterochromatic regions in this mutant. However, the aggregation of chromocenters and pericentromeric repeats suggests that heterochromatin is more condensed in *crwn1 crwn2* nuclei, despite the fact that these double mutants shared a significant portion of the mis-regulated epigenetic modifiers observed in *crwn4* mutants. This result suggests that the apparent diametrically

opposed conformation of heterochromatin in *crwn4* and *crwn1 crwn2* mutants are built upon similarly disrupted epigenetic pathways.

Another major observation in this study concerns the over-expression in *crwn* mutants of transcripts responding to biotic and abiotic stresses, featuring the cascades executed by the ERF/AP2 and WRKY transcription factors. The up-regulation profiles in *crwn* mutants covered the majority of over-expressed transcripts observed in other studies for plants subjected to a short cold treatment, a transient virus infection, or a brief mechanical stress. However, the up-regulated loci after a prolonged cold treatment, or down-regulated loci in all the studies mentioned above, did not match well with *crwn* transcriptional profiles. Therefore, *crwn* cells do not respond as if subjected tochronic external stresses. Rather, the mRNA-seq profile of *crwn* mutants exhibited a bias toward a transcriptional activation of pathways highly sensitive to acute environmental stress.

I hypothesize that *crwn* mutations mis-regulate a set of molecular switches for prompt stress response, via conformational changes of either the chromosomes or their local nuclear environments, directly affecting the transcription of corresponding loci. Future investigation should look at the conformation and positioning of the genomic loci responsible for activation of ERF/AP2 and WRKY pathways. For example, release from the nuclear periphery, where CRWN proteins are concentrated, might be involved in activation of these loci.

I also described the enrichment of cell wall and extra-cellular components in down-regulated targets in *crwn* mutants. It is known that during biotic infection, the increased production of salicylic acid (SA) is often associated with the loosening of pectin matrix [154], and down-regulation of lignin synthesis or pectin matrix production will also induce SA-mediated stress responses [155]. Consistent with this interpretation, key regulators of SA production are all up-regulated in *crwn4* and *crwn1 crwn2* mutants, including NPR1, EDS1, ALD1 and SID2 [156]. However, the activation of defense responses in *crwn* mutants does not fully explain the enrichment of cell wall and extra-cellular components among down-regulated loci. One reason is these proteins are also among up-regulated loci in response to stress, such as *Pseudomonas syringae* DC3000 and TuMV-treated plants (Supplementary Table 3.1D). Yet, these loci are not up-regulated significantly in *crwn* mutants. Also, during cold acclimation, the cell wall rearranges its structure to protect the internal membrane system, and mechanical stress usually enhances stiffness of cell walls. Both responses require extra metabolism of cell wall and extra-cellular components. Other potential mechanisms might also contribute to the enrichment of down-regulated cell wall proteins. For example, cell wall metabolism is highly active during cell division and expansion, and the lowered metabolism is consistent with the reduced DNA replication activity and endo-reduplication cycles in *crwn* mutants. It is also possible that the cell wall metabolism is directly blocked by transcriptional landscapes in *crwn* nuclei that contribute to the activation of various stress response pathways.

Finally, I focused on structural nuclear proteins to understand the mechanisms through which nuclear organization might be altered in *crwn* mutants. Despite the nuclear periphery localization of all CRWN proteins [78, 81], nuclear envelope components, including SUN-domain proteins and most nucleoporin subunits, generally retain their normal transcript expression levels in *crwn* mutants. These data suggest that these structures at the nuclear periphery are not communicating with CRWN proteins. However, I elicited both over- and under-expression of transcripts encoding SMC complex proteins in different *crwn* mutants. In yeast, mouse, and human, these complexes are involved in chromosomal condensation and sister-chromatin association in interphase, and localize to heterochromatic chromosomes during mitosis and meiosis [52]. In yeast, SMC5 and SMC6 proteins were enriched at ribosomal RNA genes and at some telomeres. *smc5* and *smc6* cells exhibit mitotic aberrations involving impaired chromosome segregation of repetitive regions [157]. The disrupted SMC5/6 complexes lead to defects in HR repair, increase DNA damage, and ectopic activation of non-homologous end joining (NHEJ) repair pathways [158]. Similarly, both SMC complexes and CRWN1 proteins localize to mitotic chromosomes in Arabidopsis.. It has also been shown that the SMC5/6 complex in Arabidopsis is important for homologous recombination-mediated double-strand break (DSB) repair via sister chromatin cohesion [102]. It is possible that the SMC complexes interact with CRWN proteins functionally and physically, and adjust their expression level in response to the loss of CRWN proteins.

Our analysis of nuclear proteins also uncovered a general down-regulation of both positive and negative regulator proteins of endo-reduplication cycles. The combination of these changes predicts reduced endopolyploidy levels especially in severe *crwn* mutants, because the down-regulation of positive regulators, such as LGO and SIM [152, 153], is more prominent. Indeed, the severe *crwn* mutants do exhibit a reduction in endopolyploidy (see Chapter 2). The hindered endo-cycles appear to be coupled with a broad down-regulation of DNA replication machinery, with very few up-regulated loci in this class in severely impaired *crwn* mutants. A reduced demand for DNA replication could be associated with a lower requirement for the deposition and maintenance of epigenetic marks on newly replicated DNA strands.

The *crwn* mutations also caused broad mis-expression of various DNA repair pathways. Several different defects in *crwn* mutants could be related to alterations in DNA repair. For example, a down-regulation of DNA replication machinery may result in more frequent replication errors. Also, the mis-expression of SMC complexes, especially SMC6A, might disrupt homologous recombination-mediated DNA repair [102], and therefore enhance the activity of a compensating repair pathway, such as base excision repair. Further, the release of transposon silencing in *crwn* mutants could create novel genomic polymorphisms (*e.g.*, DSB, copy number variations), which could induce DNA repair. In addition, the whole genome undergoes constitutive activation of defense responses, which could be coupled to a higher incidence of DNA damage [159]. These problems could concurrently slow down the progression of cell cycles and engage the cells in continuous DNA repair tasks to recover the fidelity of

genetic information. To dissect this mechanism, functional assays of DNA repair are needed. Scans to assess mutation rates in *crwn* lines should also be pursued by whole genome sequencing to detect the accumulation of DNA polymorphisms.

In summary, CRWN proteins appear to employ different mechanisms to regulate or alter gene expression. I proposed that the loss of nuclear structural CRWN proteins primarily alters nuclear organization, leading to a change in the nuclear landscape of various functional machineries (*e.g.*, transcription, DNA repair), which in turn leads to downstream effects. Specifically, I propose that the loss of CRWN proteins, especially CRWN4, expands to affect CRWN-interacting proteins, including the chromosomal scaffolding system (*e.g.*, SMC complexes), thus impairing basic nuclear processes including DNA replication, nucleosome assembly, and epigenetic modification. These defects consequentially activate DNA repair pathways to protect genomic integrity. In addition, rearrangements in nuclear organization might also alter the "transcriptional flavor" of the local environment of the chromosome, especially those regions that are co-localized with CRWN proteins on the nuclear periphery, thereby altering their transcription. The broadly activated stress response pathways might be altered in *crwn* mutants through this mechanism.

The most critical question of how the loss of CRWN proteins disrupts nuclear organization remains to be addressed. It is possible that CRWN proteins, especially CRWN4, work closely with SMC complexes, and the loss of CRWN directly down-regulates the quality and quantity of these basic components of chromosomal

organization. CRWN proteins localize to the nuclear periphery, therefore SUN-domain proteins, the only clear plant homologs of animal nuclear envelope proteins, are another candidate for CRWN-interacting proteins. Different CRWN paralogs could also form working complexes. In the next chapter, I will start to address these questions using biochemical approaches.

Materials and Methods

Plant materials and growth conditions

All T-DNA insertion alleles used in this study were obtained from the SALK collection in strain Columbia [121], and single mutant lines were originally obtained from the Arabidopsis Biological Resource Center (ABRC) at The Ohio State University. Plants were grown in long-day conditions (16 h of light / 8 h of dark) at 23˚C on soil (Metro-Mix 360, SunGro, Vancouver) in environmental growth chambers. Each genotype profiled includes three biological replicates from independent genetic lineages, and tissue from twenty individuals was pooled within each biological replicate. Mature rosette leaves were harvested from 4-week-old adult plants for each sample, and frozen in liquid nitrogen immediately for subsequent RNA purification/n.

RNA extraction and library construction

Total RNA was extracted from each biological replicate sample using the traditional Trizol® method (http://tools.lifetechnologies.com/content/sfs/manuals/trizol_reagent.pdf) and further purified using a Qiagen RNeasy kit. The total RNA sample was sent to the Genomic Core Facility at the Weill Medical School (http://corefacilities.weill.cornell.edu/genomics.html) , and the standard Illumina protocol TruSeq RNA Sample Prep Guide (15008136 A) (http://support.illumina.com/sequencing/documentation.ilmn) was used for single-stranded mRNA library construction.

Sequencing, alignment and comparison of mRNA transcriptome

18 samples from 6 genotypes were barcoded and pooled in three lanes for sequencing using an Illumina Hi seq 2000 platform. Each sample recovered 20-30 million reads 51 base pair in length. The rRNA reads were removed from the data set after alignment using Bowtie2. A Tophat-Cufflink pipeline was employed to map the reads to the genome and to perform between-sample comparisons, and a q value for each locus was calculated using the default settings in Cuffdiff 2. [160]. The DEseq pipeline in the R package was also used to conduct a parallel analysis (data not shown) to confirm that the results were consistent with the analysis that the Tophat-Cufflink pipeline produced.

Other bioinformatics analyses

Venn diagrams were generated by the online tool BioVenn (http://www.cmbi.ru.nl/cdd/biovenn/). The heatmaps were generated using HeatMapImage V6 on the public server Gene Pattern (http://genepattern.broadinstitute.org). The functional categorization was assigned to the mis-regulated loci using the gene ontology (GO) annotation tools on the TAIR website (http://www.arabidopsis.org). All other analysis was conducted using a standard spreadsheet program (Microsoft Excel).

REFERENCES

1. Egecioglu D, Brickner JH: **Gene positioning and expression**. *Current opinion in cell biology* 2011, **23**(3):338-345.
2. Kind J, van Steensel B: **Genome-nuclear lamina interactions and gene regulation**. *Current opinion in cell biology* 2010, **22**(3):320-325.
3. Dittmer TA, Stacey NJ, Sugimoto-Shirasu K, Richards EJ: **LITTLE NUCLEI genes affecting nuclear morphology in Arabidopsis thaliana**. *The Plant cell* 2007, **19**(9):2793-2803.
4. Wang H, Dittmer TA, Richards EJ: **Arabidopsis CROWDED NUCLEI (CRWN) proteins are required for nuclear size control and heterochromatin organization**. *BMC plant biology* 2013, **13**(1):200.
5. Sakamoto Y, Takagi S: **LITTLE NUCLEI 1 and 4 regulate nuclear morphology in Arabidopsis thaliana**. *Plant & cell physiology* 2013, **54**(4):622-633.
6. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shinn P, Pellegrini M, Jacobsen SE *et al*: **Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis**. *Cell* 2006, **126**(6):1189-1201.
7. Nakano T, Suzuki K, Fujimura T, Shinshi H: **Genome-wide analysis of the ERF gene family in Arabidopsis and rice**. *Plant physiology* 2006, **140**(2):411-432.

8. Eulgem T, Rushton PJ, Robatzek S, Somssich IE: **The WRKY superfamily of plant transcription factors**. *Trends in plant science* 2000, **5**(5):199-206.

9. Lee BH, Henderson DA, Zhu JK: **The Arabidopsis cold-responsive transcriptome and its regulation by ICE1**. *The Plant cell* 2005, **17**(11):3155-3175.

10. Prasch CM, Sonnewald U: **Simultaneous application of heat, drought, and virus to Arabidopsis plants reveals significant shifts in signaling networks**. *Plant physiology* 2013, **162**(4):1849-1866.

11. Walley JW, Coughlan S, Hudson ME, Covington MF, Kaspi R, Banu G, Harmer SL, Dehesh K: **Mechanical stress induces biotic and abiotic stress responses via a novel cis-element**. *PLoS genetics* 2007, **3**(10):1800-1812.

12. Thilmony R, Underwood W, He SY: **Genome-wide transcriptional analysis of the Arabidopsis thaliana interaction with the plant pathogen Pseudomonas syringae pv. tomato DC3000 and the human pathogen Escherichia coli O157:H7**. *The Plant journal : for cell and molecular biology* 2006, **46**(1):34-53.

13. Kohorn BD: **WAKs; cell wall associated kinases**. *Current opinion in cell biology* 2001, **13**(5):529-533.

14. Schubert V: **SMC proteins and their multiple functions in higher plants**. *Cytogenetic and genome research* 2009, **124**(3-4):202-214.

15. Watanabe K, Pacher M, Dukowic S, Schubert V, Puchta H, Schubert I: **The STRUCTURAL MAINTENANCE OF CHROMOSOMES 5/6 complex promotes sister chromatid alignment and homologous recombination after**

DNA damage in Arabidopsis thaliana. *The Plant cell* 2009, **21**(9):2688-2699.

16. Roeder AH, Chickarmane V, Cunha A, Obara B, Manjunath BS, Meyerowitz EM: **Variability in the control of cell division underlies sepal epidermal patterning in Arabidopsis thaliana**. *PLoS biology* 2010, **8**(5):e1000367.
17. Churchman ML, Brown ML, Kato N, Kirik V, Hulskamp M, Inze D, De Veylder L, Walker JD, Zheng Z, Oppenheimer DG *et al*: **SIAMESE, a plant-specific cell cycle regulator, controls endoreplication onset in Arabidopsis thaliana**. *The Plant cell* 2006, **18**(11):3145-3157.
18. Bethke G, Grundman RE, Sreekanta S, Truman W, Katagiri F, Glazebrook J: **Arabidopsis PECTIN METHYLESTERASEs Contribute to Immunity against Pseudomonas syringae**. *Plant physiology* 2014, **164**(2):1093-1107.
19. Gallego-Giraldo L, Escamilla-Trevino L, Jackson LA, Dixon RA: **Salicylic acid mediates the reduced growth of lignin down-regulated plants**. *Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**(51):20814-20819.
20. Ng G, Seabolt S, Zhang C, Salimian S, Watkins TA, Lu H: **Genetic dissection of salicylic acid-mediated defense signaling networks in Arabidopsis**. *Genetics* 2011, **189**(3):851-859.
21. Hirano T: **At the heart of the chromosome: SMC proteins in action**. *Nature reviews Molecular cell biology* 2006, **7**(5):311-322.

22. Torres-Rosell J, Machin F, Farmer S, Jarmuz A, Eydmann T, Dalgaard JZ, Aragon L: **SMC5 and SMC6 genes are required for the segregation of repetitive chromosome regions**. *Nature cell biology* 2005, **7**(4):412-419.

23. Kegel A, Sjogren C: **The Smc5/6 complex: more than repair?** *Cold Spring Harbor symposia on quantitative biology* 2010, **75**:179-187.

24. Yan S, Wang W, Marques J, Mohan R, Saleh A, Durrant WE, Song J, Dong X: **Salicylic Acid activates DNA damage responses to potentiate plant immunity**. *Molecular cell* 2013, **52**(4):602-610.

25. Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R *et al*: **Genome-wide insertional mutagenesis of Arabidopsis thaliana**. *Science* 2003, **301**(5633):653-657.

26. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: **TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions**. *Genome biology* 2013, **14**(4):R36.

27. Woo HR, Dittmer TA, Richards EJ: **Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis**. *PLoS genetics* 2008, **4**(8):e1000156.

28. Stroud H, Greenberg MV, Feng S, Bernatavichute YV, Jacobsen SE: **Comprehensive analysis of silencing mutants reveals complex regulation of the Arabidopsis methylome**. *Cell* 2013, **152**(1-2):352-364.

29. Lister R, O'Malley RC, Tonti-Filippini J, Gregory BD, Berry CC, Millar AH, Ecker JR: **Highly integrated single-base resolution maps of the epigenome in Arabidopsis**. *Cell* 2008, **133**(3):523-536.

CHAPTER FOUR

THE NUCLEAR COILED-COIL PROTEINS CRWN1 AND CRWN4 PHYSICALLY INTERACT TO REGULATE NUCLEAR ORGANIZATION IN ARABIDOPSIS THALIANA

Abstract

In this study I describe the biochemical characterization of the nuclear coiled-coil proteins CRWN1 and CRWN4, which are required for proper maintenance of nuclear organization in the flowering plant *Arabidopsis thaliana*. Polyclonal antisera against CRWN1 and CRWN4 were used as immunological probes to study the localization and interaction of these proteins. I demonstrate that both CRWN1 and CRWN4 are enriched in nuclear extracts and are resistant to salt extraction. Through a co-immunoprecipitation assay, I provide evidence of a physical interaction between CRWN1 and CRWN4 proteins, supporting the hypothesis that CRWN1 and CRWN4 form functional complexes *in vivo*. The abundance of CRWN4 is significantly reduced in *crwn1* mutant backgrounds, while a complex feedback regulation on the mRNA level exist among different *CRWN* genes. My results suggest that a balance between CRWN1-like and CRWN4-like proteins is important to accomplish the proper regulation of nuclear organization.

Introduction

Nuclei are the cellular settings for essential functions, such as DNA replication and repair, epigenetic modification, and transcription. Dynamic nuclear organization plays an important role in regulating these molecular activities [45]. An important regulator of nuclear organization is the nuclear lamina, a lattice-like structure underlying the inner nuclear envelope in animal cells composed of the type V intermediate filament protein, lamin, and its interacting proteins (*i.e.*, 'lamin-associated proteins') [45]. Mutations in the human genes encoding lamin A/C, *LMNA*, and lamin-associated proteins cause a complex set of clinical syndromes called laminopathies [58]. The nuclei of cells from laminopathy patients are irregular looking and fragile, and contain redistributed histone modification marks [45, 58].

Lamins have been reported to exist only in metazoans, although NUP1 in Trypanosomes [60] and NE81 in Dictyostelium have been proposed to be candidate lamin analogs [161]. There are no plant homologs resembling lamin proteins [78], and it is unclear how plants organize their nuclei. In the 1990s, Masuda and colleagues discovered the coiled-coil domain protein NMCP1 (Nuclear Matrix Constituent Protein 1), using a monoclonal antibody screen for antigens in nucleoskeleton preparations from carrot cells [75]. Since then, NMCP1 and its paralogs have been considered to be prime candidates for the central components of a plant nuclear lamina [78, 79, 120, 162] .

NMCP proteins harbor nuclear localization signals and a variety of cytological observations, including immunolocalization and fluorescent protein tagging experiments, demonstrate the NMCPs are nuclear proteins [78, 79, 81, 115]. The Arabidopsis NMCP proteins, called CRWN [140] (and previously known as LINC [78]), also contain long coiled-coil domains in their central regions, and are localized primarily at the nuclear periphery. Our previous studies showed that CRWN1 and CRWN2 proteins concentrate at the nuclear periphery, but over-expression of these two proteins, particularly CRWN2, led to their distribution throughout the nucleoplasm [78]. Localization of this class of protein at the nuclear periphery is supported by electron microscope immunostaining experiments in onion (*Allium cepa*) that show AcNMCP1 proteins were concentrated around the nuclear rim, with some speckled signals in the nucleoplasm [77]. A recent study also reported that LINC4 (CRWN4) proteins are located at the nuclear periphery, and further showed that LINC1 proteins co-localize with mitotic chromosomes during the cell cycle [81]. Similarly, AgNMCP1 in celery (*Apium graveolens*) co-localizes with the mitotic spindle and segregating chromosomes, while AgNMCP2 is dispersed throughout the mitotic cytoplasm. These two proteins accumulate at the nuclear periphery in anaphase at the end of the cell cycle [76].

NMCP proteins are characterized by insolubility during biochemical treatment of isolated nuclei. Both the carrot (*Daucus carota*) prototypic protein DcNMCP1 and its onion (*Allium cepa*) ortholog AcNMCP1 fractionate with insoluble nuclear extracts

resistant to salt and non-ionic detergent extraction [75, 79]. This behavior is not simply associated with the hydrophobicity typical of nuclear membrane proteins, because NMCP proteins do not possess any conventional transmembrane domains. An immunoprecipitation-based proteomic study of GFP-tagged nucleoporins in Arabidopsis failed to detect any NMCP family proteins interacting with the several tested nucleoporins [134]. It is more likely that NMCP proteins connect to the inner nuclear membrane via other interacting partners, and the biochemical fractionation characteristics could be due to the long coiled-coil domain in the center of NMCP proteins. The coiled-coil structure is a widespread motif involved in oligomerization and interaction with DNA or other proteins [163]. One classical example of a single coiled-coil is the leucine zipper structural motif in various transcription factors, which consist of parallel, left-handed homodimers for DNA binding [164]. It is common for proteins containing long coiled-coil domains to polymerize with each other to form insoluble networks and serve structural functions in cells. Examples include SMC (Structural Maintenance of Chromosomes) proteins [100], tropomyosins [53], and lamin proteins [48, 51].

In this study, I investigate biochemical aspects of CRWN proteins. My results show that CRWN1 and CRWN4 proteins fractionated with the insoluble nuclear residue, which was resistant to high salt and non-ionic detergent treatment. *CRWN4* mRNA was up-regulated in *crwn1* mutants; however, CRWN4 protein abundance was greatly reduced in the absence of CRWN1. Moreover, CRWN1 and CRWN4 proteins interacted with each other *in vivo*. The complex regulation among *CRWN* genes on

both the transcriptional and protein levels supports a model wherein the balance between CRWN1-like and CRWN4 functions are important for the regulation of nuclear organization.

Results

The structure and natural variation of CRWN family proteins

Figure 3.1 depicts the structure of the Arabidopsis CRWN proteins. Each protein is approximately 1100 amino acids in length with a long coiled-coil domain starting from the second or third exon and extending to the middle of the largest sixth exon (black bar). In addition, CRWN1, 2 and 3 contain a well-conserved C-terminus (purple bar), which is absent in CRWN4. Polymorphic amino acid sites of CRWN proteins within species *Arabidopsis thaliana* were identified by comparing predicted protein sequences of various natural accessions from the 1001 Genome Project [165, 166]; Figure 4.1 shows the distribution of sites exhibiting low and high levels of polymorphisms (Figure 4.1). Close to 10% of the amino acid residues of each CRWN protein were identified as polymorphic sites (Table 4.1), and they are randomly distributed across the protein, with no obvious clustering in any particular domain. Polymorphism levels are similar among the four CRWN proteins (Table 4.1), and polymorphic patterns among the four CRWN proteins were not correlated.

To broaden the scope of this investigation of natural variation, I extended the comparison beyond *Arabidopsis thaliana* to its close relative *Arabidopsis lyrata.* There are four CRWN paralogs in *A. lyrata*, corresponding to CRWN1, CRWN2, CRWN3 and CRWN4 in *A. thaliana*. Specifically, I compared the CRWN protein sequences of the *A. lyrata* genome (version 1.0, one genotype) with the consensus sequence of the corresponding CRWN proteins in *A. thaliana*. A few insertions and deletions were detected (pink gaps) between the *A. thaliana* and *A. lyrata* proteins, and dozens of sites (light green) in *A. lyrata* differed from those seen in *A. thaliana.* The majority of these variations were *A. lyrata*-specific, suggesting independent evolutionary forces are exerted on these two species, or that the variation occurred randomly after divergence of a common ancestor. Among these *A. lyrata-A.thaliana* polymorphisms, the abundance of these sites in the AtCRWN4 protein was the lowest among the four CRWN paralogs. This pattern contradicts the typical situation in which the divergence patterns for different gene families between *A. lyrata* and *A. thaliana* mirror within-*A thaliana* polymorphism levels [167]. It is possible that a selective pressure in *A. lyrata* constrained the accumulation of polymorphisms in the CRWN4 protein among natural strains, while the presence of three copies of CRWN1-like proteins [140] releases that pressure through redundancy. Nonetheless, due to the fact that only one *A. lyrata* genome was available for comparison, it is difficult to assess the significance of these polymorphism patterns.

Figure 1

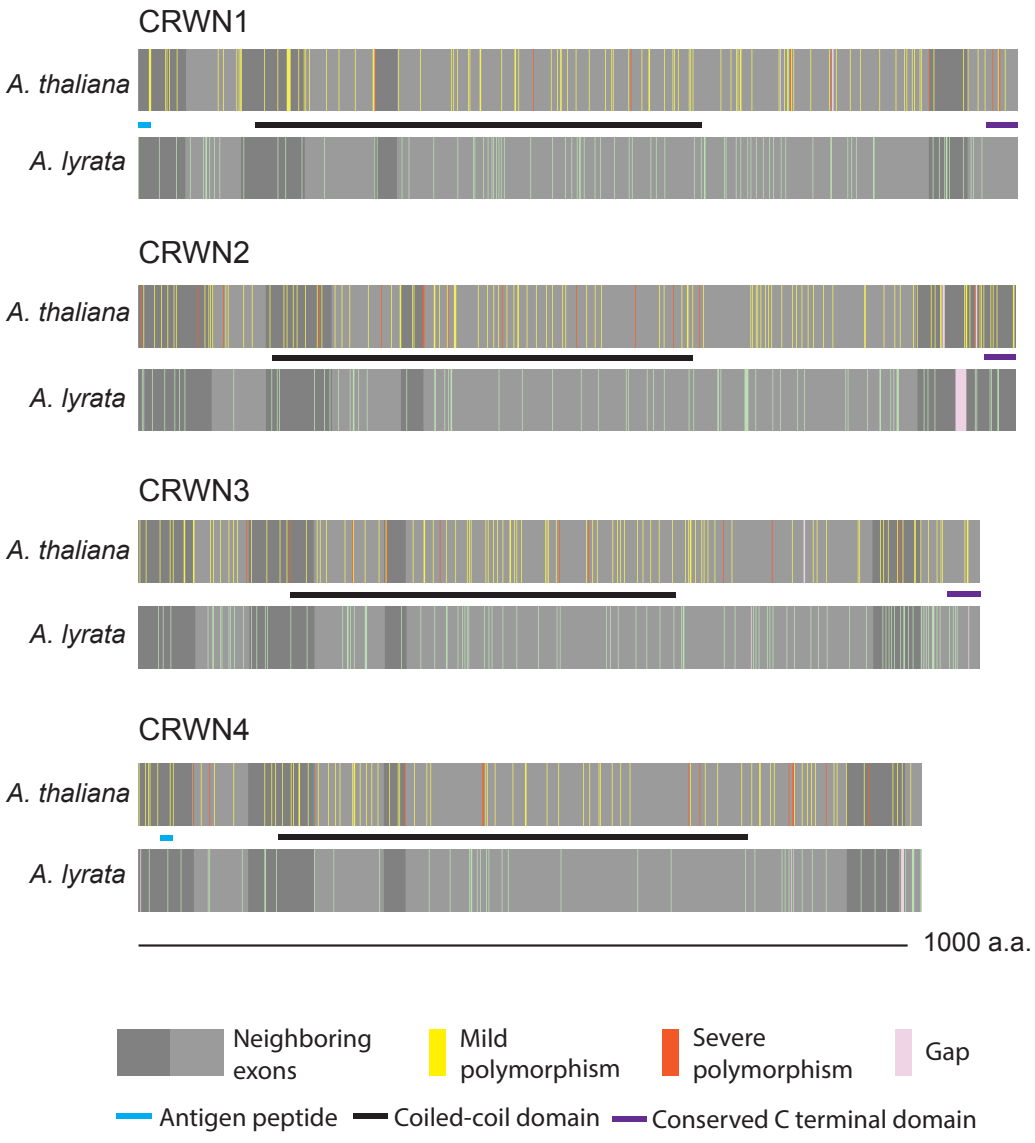


Figure 4.1 CRWN protein domains and sequence polymorphism

Figure 4.1 The domain composition of CRWN proteins is illustrated. The amino acid sequences of CRWN proteins are displayed in gray boxes, and the neighboring exons are indicated by alternated light and dark gray colors. The conserved C-termini, predicted long coiled-coil domain [78], and N-terminal sequences used for antigen design, are shown by separate bars in purple, black, and blue, respectively, for each CRWN protein. Polymorphic sites within *Arabidopsis thaliana* are displayed in yellow (mild polymorphisms with similar biochemical properties) or red (severe polymorphisms with distinct biochemical properties) determined using a Gonnet PAM 250 matrix with a cut off value of 0.5 [168]. The predicted CRWN protein sequences from the *Arabidopsis lyrata* 1.0 genome were displayed in a similar manner, with polymorphic sites divergent from the *A. thaliana* consensus sequence colored in light green. Pink gaps indicate the insertions and deletions in the alignments between the two species.

Arabidopsis thaliana 1001 genome

polymorphism	severe	mild	total	protein size	%
CRWN1	8	96	104	1132	9.19%
CRWN2	14	101	115	1128	10.20%
CRWN3	10	94	104	1085	9.59%
CRWN4	13	73	86	1010	8.51%

A lyrata v.s. A thaliana

polymorphism	insertion	deletion	changes	total	protein size	%
CRWN1	3	0	81	84	1135	7.40%
CRWN2	5	14	73	92	1119	8.22%
CRWN3	2	2	93	97	1085	8.94%
CRWN4	1	6	44	51	1005	5.07%

overlap with thaliana polymorphic sites

polymorphism	lyrata only	severe	mild
CRWN1	77	2	5
CRWN2	81	1	10
CRWN3	77	5	15
CRWN4	44	5	2

Table 4.1 Summary of CRWN protein sequence polymorphisms

Table 4.1 The statistics for Figure 4.1. Numbers of polymorphic sites in each CRWN protein within *A. thaliana,* and between *A. thaliana* and *A. lyrata* are counted and compared.

Development of antibody probes recognizing CRWN1 and CRWN4

My previous genetic analysis demonstrated that CRWN1 and CRWN4 are the two major proteins in the CRWN family, possessing phylogenetically distinct amino acid sequences and exhibiting different nuclear functions (see Chapter 2). However, the loss of CRWN1-like and CRWN4 functions mis-regulates common loci at the transcription level (see Chapter 3), suggesting that CRWN1 and CRWN4 might work together to fulfill their apparently divergent functions. Thus, I focused on these proteins as representatives of this protein family. To develop immunological probes to study CRWN function, peptide antigens were designed to raise antisera specific for either CRWN1 or CRWN4. Two ~20 amino acid sequences specific for the targeted proteins were chosen from the N-termini of the protein sequences (blue bar in Figure 4.1) for three reasons. First, terminal sequences are more likely to be exposed on the surface of well-folded proteins *in vivo* [169]; therefore, these regions might be more readily available for antibody binding. Second, the C-termini of CRWN family proteins are relatively conserved compared to the N-termini [78], suggesting that the C-terminus might have more important functional constraints, for example, by participating in protein-protein interactions. Consequently, I focused on a less conserved region for antisera recognition and binding in hopes of avoiding disruption of interactions among CRWN proteins and any relevant working partners. Third, the N-termini of CRWN proteins are more variable, thus it is more likely to find a region specific to either CRWN1 or CRWN4 proteins.

The peptide antigens were synthesized, injected into rabbits, and corresponding antisera were affinity purified (see Materials and Methods). I screened antisera raised against CRWN1 or CRWN4 peptides using western blots of nuclear extracts from seedlings of wild type and corresponding *crwn* mutants, and I identified specific antisera for the two CRWN proteins (Figure 4.2A). The antibody against CRWN1 detected a series of four distinct bands between 100 kDa and 150 kDa, while no signal was detected in *crwn1* mutant samples, indicating that all the bands detected by this antibody are specific to CRWN1. Among the four bands, the brightest signal shifted from the larger bands toward the smaller 110 kDa band as extraction time was extended (Figure 4.2B). The 110 kDa band was dominant and reproducible compared to other bands across all our experiments (data not shown), and the 110 kDa size is consistent with the molecular weight of CRWN1 protein. Thus, I used the 110 kDa as a signature of the presence of CRWN1 proteins in our studies. The antibody against CRWN4 protein detected three equally bright bands ranging in size from 110 kDa to 150 kDa in wild type samples. All of these bands were missing from *crwn4* mutant samples, indicating that these three bands specifically recognize CRWN4 proteins. It is likely that the distinct bands detected by the antisera represent different isoforms and/or modifications of these two proteins *in vivo*. Additional characterization, for example, using mass spectrometry, will be necessary to determine the full spectrum of the protein isoforms and modifications. With these validated immunological probes in hand, I turned my attention to the following biochemical investigations.

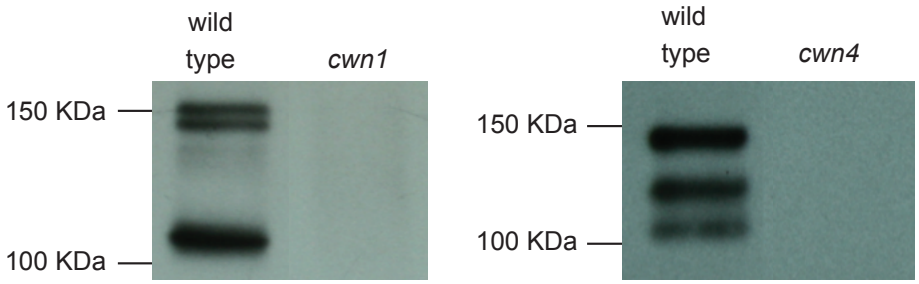


Figure 4.2 Specificity test for antisera against CRWN1 and CRWN4 proteins

Figure 4.2 Detection of CRWN1- and CRWN4- specific bands using antisera raised for this study. Proteins from nuclear extracts of Col wild type, *crwn1,* and *crwn4* seedlings were resolved by SDS-PAGE protein electrophoresis, and proteins were detected using western immunoblot techniques.

CRWN1 and CRWN4 fractionate with the salt-resistant insoluble nucleoskeleton

I first investigated the sub-nuclear fractionation characteristics of CRWN1 and CRWN4. Many investigators have used high salt and detergent to perform nuclear extractions to generate an insoluble 'nuclear matrix' fraction resistant to these harsh biochemical conditions [170]. This residual framework is thought to represent the network of fibers that exist throughout the nucleus. To test the hypothesis that CRWN proteins are structural components in plant nuclei, I followed the above criteria and applied various salt and detergent treatments to Arabidopsis nuclear extracts to solubilize proteins, and determined whether CRWN proteins remained in the insoluble pellet after separation by centrifugation.

Crude nuclear preparations from seedling tissue were extracted using various concentrations of NaCl. The results displayed in Figure 4.3A show a control chromatin protein, histone H2B, was solubilized to a significant extent by NaCl treatment at a concentration of 0.5M and higher. In contrast, only faint CRWN1 bands were recovered in the salt-treated supernatant samples, while strong signals were detected in all pellet samples. This partitioning indicates that most of the CRWN1 protein remained insoluble after extraction, resulting in a pattern resembling that seen in the control sample to which no salt was applied. The salt treatment only solubilized a very small portion of CRWN1 proteins regardless of the NaCl concentration in the buffer. Similarly, no significant CRWN4 signal was detected among the solubilized

samples; the three CRWN4 bands were recovered in the pellet samples with equal signal strength, indicating that CRWN4 is also resistant to salt extraction.

In addition to salt treatment, various concentrations of detergents, including Triton X-100, CHAPS, and SDS, were applied to assess the solubility of CRWN proteins. Triton X-100 by itself did not affect the solubility of CRWN1 or CRWN4 (Figure 4.3A, Supplementary Figure 4.1A). Different concentrations of CHAPS also did not significantly affect CRWN1 or CRWN4 solubility (Figure 4.3B), although in some cases, a low concentration of CHAPS slightly solubilized CRWN4 (data not shown). In contrast, SDS efficiently solubilized both CRWN1 and CRWN4 (Supplementary Figure 4.1), even at a low (0.2%) concentration. Nonetheless, a portion of the proteins remained in the nuclear pellet after extraction with SDS (Supplementary Figure 4.1). These data demonstrate that CRWN proteins were resistant to high salt and mild detergent extraction, suggesting that CRWN1 and CRWN4 are components of the nuclear matrix.

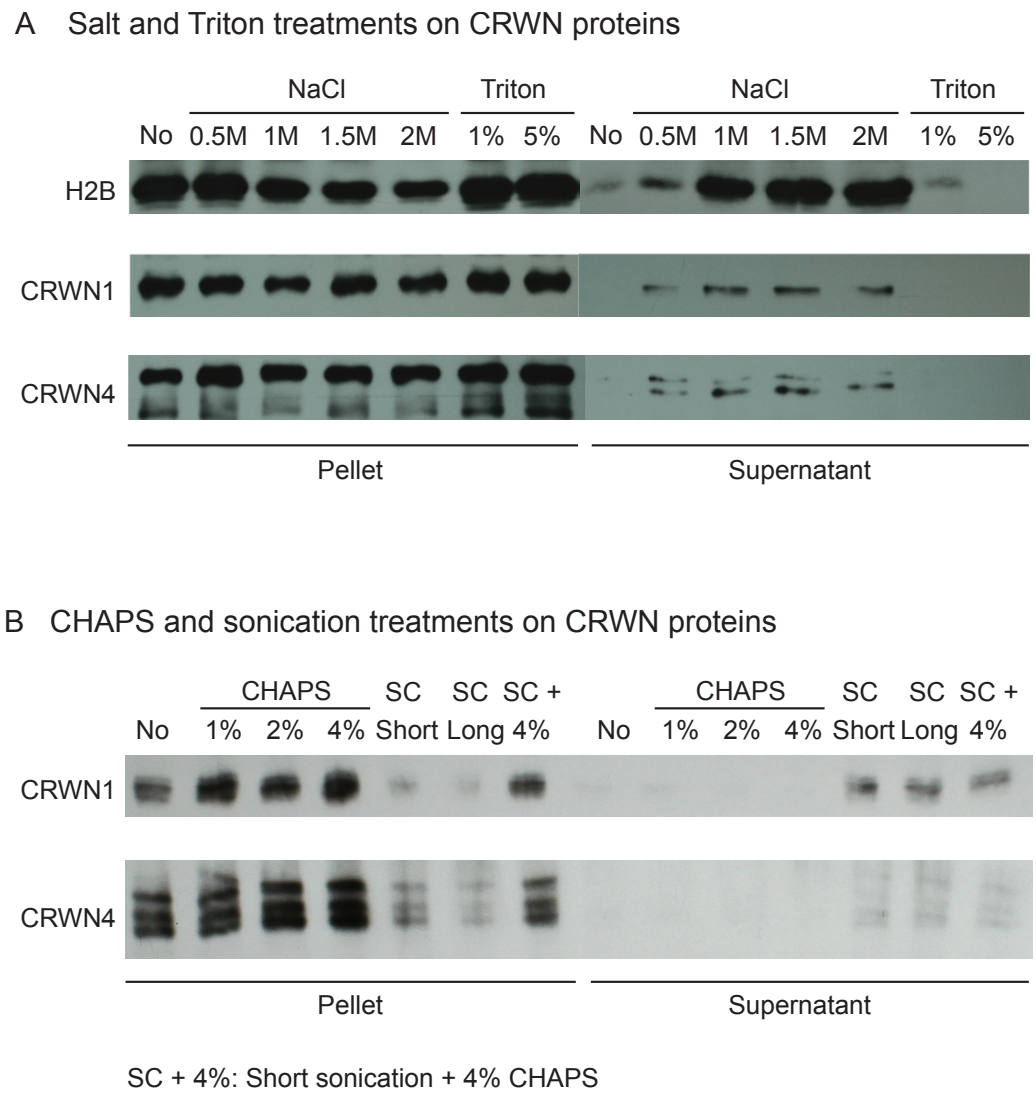
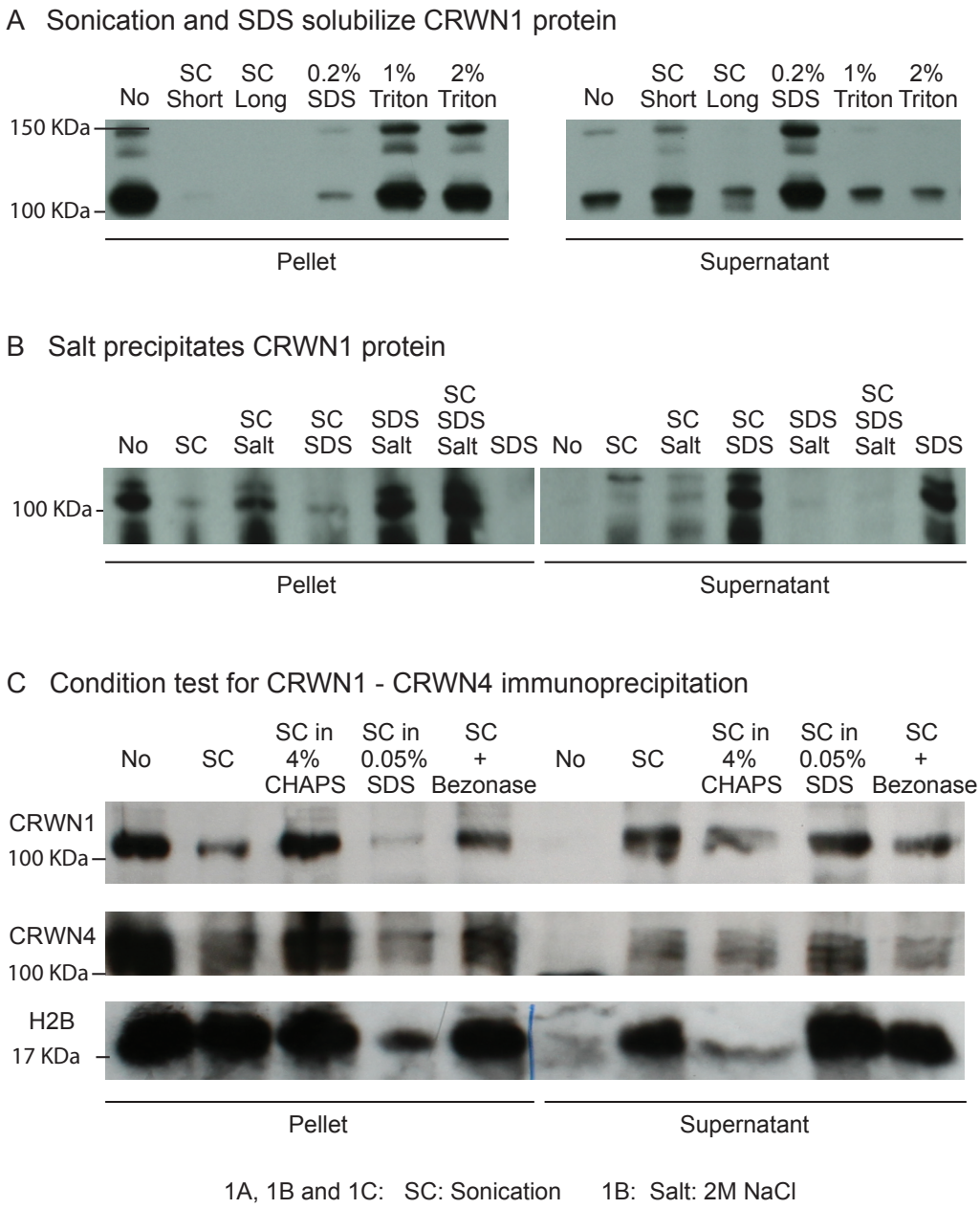


Figure 4.3 CRWN1 and CRWN4 proteins are insoluble under mild detergent and high salt extraction conditions

Figure 4.3 Detection of CRWN1 and CRWN4 proteins in salt- or detergent- treated nuclear extracts of wild type samples. A concentration series of NaCl (0M, 0.5M, 1M, 1.5M, 2M), Triton X-100 (1%, 2%) and CHAPS (1%, 2%, 4%) were applied. The left panel represents the insoluble pellets, while the right panel represents the solubilized supernatants. The signal against CRWN1 proteins (the major band is a little bit above 100 kDa) was displayed in the upper panel, and CRWN4 proteins (triplet bands between 100 -150 kDa) in the bottom panel. Histone H2B (17 kDa) was used as loading control. SC indicates sonication treatment.



Supplementary Figure 4.1 The optimization of conditions for CRWN1-CRWN4 co-immunoprecipitation

Supplementary Figure 4.1 Detection of CRWN1 and CRWN4 proteins in sonication, salt-, and detergent-treated nuclear extracts of wild type samples. In Panel A, B and C, Short (5 sec x 10 times) and long (5 sec x 20 times) sonications, NaCl (2M), Triton X-100 (1%, 2%), CHAPS (4%), SDS (0.05%, 0.2%) and Benzonase ® (Sigma-Aldrich) were applied individually or in combination. The left panel represented the insoluble pellet, while the right panel represented the solubilized supernatants. The signal against CRWN1 proteins (the major band is a little bit above 100 kDa) and CRWN4 proteins (triplet bands between 100 -150 kDa) were displayed. In Panel C, histone H2B (17 kDa) was used as loading control.

The abundance of CRWN4 protein is dependent on CRWN1 protein expression

Phylogenetic analyses place CRWN family proteins into two major clades [140]: the CRWN1-like clade (CRWN1, 2 and 3 in Arabidopsis) and the CRWN4 clade. The genetic analysis presented in Chapter 2 further supports this classification by showing non-redundant functions for CRWN1-like and CRWN4 proteins in controlling whole-plant phenotypes and nuclear morphology. However, my mRNA-seq analysis (see Chapter 3) demonstrate that despite the structural and functional divergence between CRWN1-like and CRWN4 proteins, these proteins share common genomic targets in transcriptional regulation. Based on these observations, I propose that CRWN1 and CRWN4 work coordinately to carry out their functions. I further hypothesize that the cooperation of CRWN1 and CRWN4 occurs because the proteins physically interact in a functional complex within the nucleus.

To test this hypothesis, I first used protein immunoblots to examine whether the abundance of CRWN1 or CRWN4 protein was altered in *crwn* single and double mutants. Figure 4.4 displays the distinct bands of the CRWN1 protein pool in wild type, *crwn2, crwn3,* and *crwn4* single mutants, as well as *crwn2 crwn3*, *crwn2 crwn4*, and *crwn3 crwn4* double mutants, in which the *crwn1* mutation was not present. Interestingly, the triplet bands representing CRWN4 proteins were not only absent from the *crwn4* single mutant, but were significantly reduced in the *crwn1* single mutant. Among the double mutants, *crwn2 crwn3* was the only genotype exhibiting a strong signal for the CRWN4 triplet bands. These signals were severely diminished in

crwn1 crwn2 and *crwn1 crwn3* samples, and not detectable in the *crwn1 crwn4, crwn2 crwn4*, and *crwn3 crwn4* mutant backgrounds. Any plant that carried a *crwn1* mutation lost the CRWN1 protein and suffered a dramatic reduction in the abundance of the CRWN4 protein. However, the loss of CRWN4 did not have a significant effect on the abundance of CRWN1. These results suggest that the presence of CRWN1 is required for the proper production or stability of CRWN4.

These observations prompted me to look in my mRNA-seq dataset at the abundance of *CRWN* mRNA in different *crwn* mutants, especially *CRWN4* transcripts in *crwn1* single mutants. Figure 4.5 summarizes the average RPKM (Reads Per Kilobase of exon per Million fragments mapped) values of the mRNA for the four *CRWN* genes and a control gene, cyclophilin. The abundance of *CRWN4* transcripts was elevated approximately 30% in *crwn1* mutants, indicating that loss of CRWN1 leads to a modest boost in *CRWN4* expression at the mRNA level. Therefore, down-regulation at the transcript level cannot explain the significant reduction in CRWN4 protein expression I observed in Figure 4.4. Rather, this observation suggests that a compensatory feedback up-regulation of *CRWN4* mRNA occurs in response to either the loss of CRWN1 or the reduction in CRWN4 protein abundance.

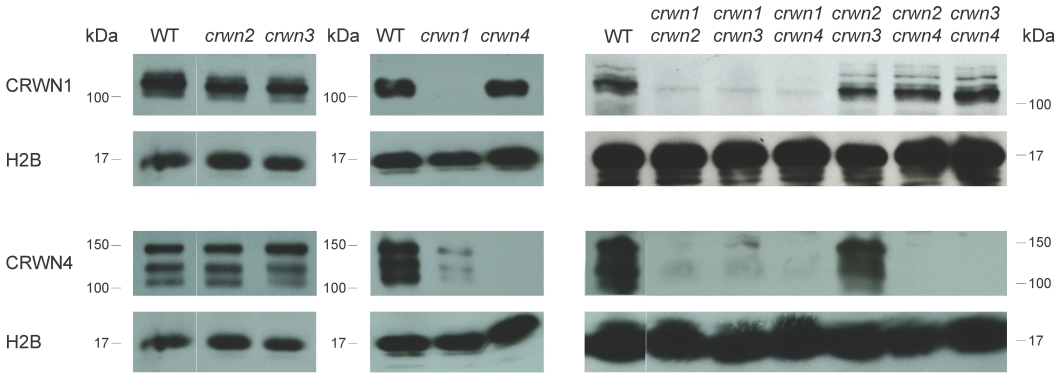


Figure 4.4 CRWN4 abundance is reduced in crwn1 backgrounds

Regulation of *CRWN* expression at the mRNA level

Figure 4.4 Detection of CRWN1 and CRWN4 proteins in *crwn* single and double mutants using protein immunoblots. The signals against CRWN1 protein (the major band is slightly above 100 kDa) were displayed in the upper panel, and the signals against CRWN4 proteins (triplet bands between 100 – 150 kDa) in the bottom panel. Histone H2B (17kDa) was used as loading control. In each of the three columns, all lanes shown are from the same experiment. Each row within a column is from the same film exposure, but some lanes were rearranged for clarity.

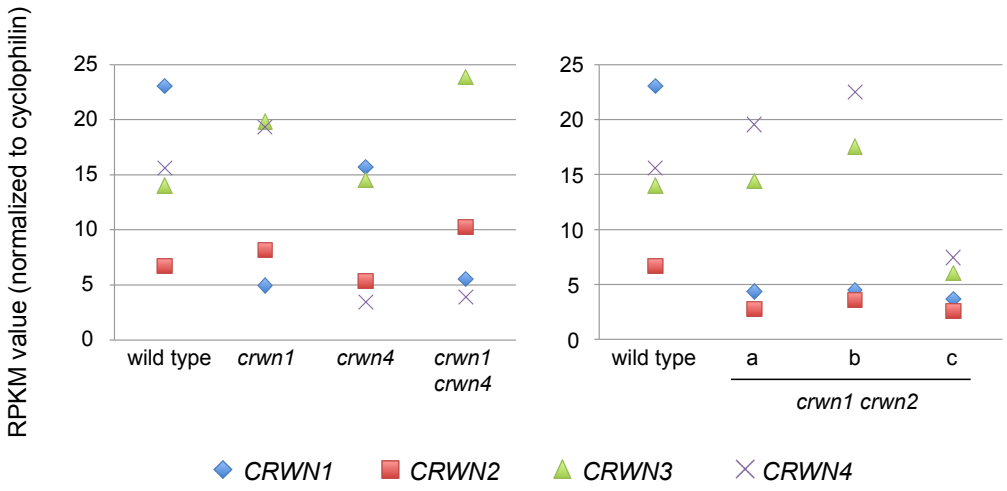


Figure 4.5 The expression of CRWN mRNAs in various crwn mutants

Figure 4.5 The left panel shows the normalized average RPKM value of each *CRWN* mRNA in wild type, *crwn1, crwn4*, and *crwn1 crwn4* mutants. The right panel showed the normalized RPKM value of each replicate in the *crwn1 crwn2* mutant sample in comparison to wild type. The RPKM value of the cyclophilin locus was used as internal control for normalization.

Analysis of my mRNA-seq datasets for *crwn* mutants also revealed a complicated regulatory relationship among *CRWN* transcripts. First, I observed an up-regulation of *CRWN2* and *CRWN3* mRNA in *crwn1* mutant backgrounds, suggesting that a transcriptional compensation occurs among *CRWN1*-like genes. Second, *CRWN1* mRNA was down-regulated in *crwn4* mutants, although a significant change in CRWN1 protein abundance was not observed by protein immunoblots. Nonetheless, the down-regulation of *CRWN1* transcripts in *crwn4* mutants indicates that the presence of the CRWN4 protein is required for normal production of transcripts encoding CRWN1-like proteins.. Third, in *crwn1 crwn4* double mutants, *CRWN2* and *CRWN3* transcripts were expressed at a higher level, especially *CRWN3*, which nearly doubles in abundance. This elevation of *CRWN2* and *CRWN3* transcripts might be responsible for the apparent functional suppression between *crwn1* and *crwn4* mutations. Taken together, these observations highlight the complex interactions and regulation among different *CRWN* genes on the mRNA level, and aides the interpretation of the phenotypes of plants carrying different *crwn* mutations. In addition, these interactions further support the hypothesis that the CRWN family of coiled-coil proteins function together on the biochemical level.

CRWN1 and CRWN4 proteins function in the same complexes

To test this hypothesis, I used immunoprecipitation to examine whether CRWN1 and CRWN4 proteins physically interact with each other *in vivo*. There were several technical challenges I needed to overcome for this experiment, and I began by

adapting an established protocol for Arabidopsis protein co-immunoprecipitation [171]. First, to concentrate CRWN proteins, nuclear extracts from large amounts of seedlings tissues were used. Public microarray database compilations (eFP browser; http://bar.utoronto.ca/efp/cgi-bin/efpWeb.cgi) showed that the *CRWN* mRNA accumulation is highest among proliferating tissues (*e.g.*, root tips and shoot meristems). Further, our lab's unpublished translational fusion reporter data shows that CRWN1 expression at the protein level is concentrated in meristematic regions and that the expression drops dramatically in mature tissues.

A second obstacle was the insolubility of CRWN proteins (Figure 4.3). Efficient solubilization with strong ionic detergents, such as SDS, might disrupt native protein-protein interactions. To maximize the recovery of potential CRWN complexes from the nuclear extracts, lysis conditions needed to be optimized to achieve a balance between solubilizing CRWN complexes and preserving their native interactions. As shown in Supplementary Figure 4.1A-C, different conditions were applied to wild type nuclear extracts, and the solubility of CRWN proteins were assessed by western blotting. In Panel 1A, a short sonication in a buffer containing 0.2% SDS solubilized CRWN1 proteins well, with weak bands left in pellet samples and strong signals present in supernatant samples. Longer sonication treatment resulted in a loss of CRWN1 signals in both pellet and supernatant samples, indicating protein degradation. Panel 1B confirmed that sonication solubilized CRWN1 proteins but also led to degradation; while SDS treatment by itself or in combination with sonication efficiently solubilized CRWN1 proteins. However, high salt (2M NaCl) treatment, by

itself or in combination with either SDS or sonication, did not solubilize CRWN1 proteins. This result indicated that high salt precipitates rather than solubilizes CRWN1 proteins, even in the presence of SDS detergent and sonication. Based on this information, I tested the conditions for CRWN1-CRWN4 co-immunoprecipitation. As shown in Supplementary Figure 4.1 Panel 1C, I piloted a sonication approach, in combination with detergent treatment (*e.g.*, CHAPS (4%), low concentration SDS (0.05%)) or application of the chromatin depletion enzyme, Bezonase [171]. Ultimately, I determined that the 0.05% SDS + sonication treatment solubilized both CRWN1 and CRWN4 best, and these conditions were used in the following investigations.

A third concern was that the antisera I raised might not have enough specificity and binding affinity to recognize and pull out the CRWN complexes from the protein extracts in the relatively harsh immunoprecipitation buffer conditions. To maximize the yield, protein extracts were incubated with antisera thoroughly, before the antibody affinity beads were added for the recovery. In addition, I used antibody affinity beads with higher specificity (*i.e.*, Dynabeads versus Agrobeads, see Materials and Methods) to reduce the loss of recovered complexes during the washing steps.

Using these modifications, I conducted reciprocal co-immunoprecipitation experiments in wild type, *crwn1* and *crwn4* mutant samples. Figure 4.6A shows the detection of CRWN1 protein. In wild type samples, as expected, CRWN1 bands were present in the anti-CRWN1 precipitate pellets but absent from the no-antiserum

control pellets. A strong signal of CRWN1 protein was detected in the anti-CRWN4 pellets, indicating that CRWN1 protein could be precipitated by anti-CRWN4 antisera. The same sets of experiments were performed in the *crwn1* and *crwn4* mutant backgrounds. As expected, no CRWN1 signals were present in the *crwn1* mutant sample pellets. In the *crwn4* background, CRWN1 protein was still present in the anti-CRWN1 precipitates, but absent from the anti-CRWN4 precipitates, suggesting that CRWN1-CRWN4 complexes no longer exist due to the loss of CRWN4 protein. Panel B of Figure 4.6 shows the detection of the CRWN4 protein on the same membrane. In the wild type samples, CRWN4 bands were detected in the anti-CRWN4 precipitates but absent from the no-antiserum control. At the same time, a faint signal of the CRWN4 triplet bands were present in the anti-CRWN1 pellets, suggesting that CRWN4 protein could be precipitated by anti-CRWN1 antisera, although at a lower efficiency. This weak binding from the reciprocal co-immunoprecipitation has been replicated several times (data not shown). In *crwn4* mutants, no CRWN4 signal was detected as expected. In the *crwn1* background, CRWN4 protein was absent from not only the anti-CRWN1 pellets but also from the anti-CRWN4 precipitates, indicating that the abundance of CRWN4 protein in the *crwn1* background is too low for it to be precipitated or detected. These results demonstrate that CRWN1 and CRWN4 proteins are components of the same complex(es) *in vivo*.

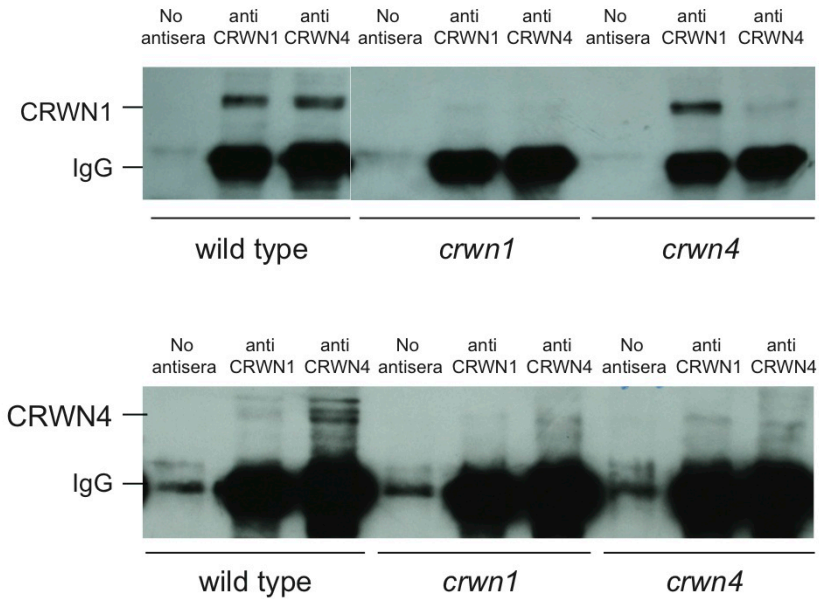


Figure 4.6 CRWN1 and CRWN4 proteins interact with each other in vivo

Figure 4.6 Detection of CRWN1 protein (upper panel) and CRWN4 protein (bottom panel) in three parallel immunoprecipitation experiments (no antisera, anti-CRWN1 antisera, and anti-CRWN4 antisera) are displayed for wild type, *crwn1*, and *crwn4* mutants. The 110kDa band represents CRWN1 proteins, and the 100kDa-150kDa triplet bands correspond to CRWN4 proteins. All lanes shown are from the same experiment and the same film exposure; but some lanes were rearranged for clarity.

Discussion

In this chapter I undertook a biochemical characterization of the putative nuclear architectural proteins, CRWN1 and CRWN4. First, I developed native antisera against CRWN1 and CRWN4 proteins. I fractionated Arabidopsis cells and used western blots to reveal that CRWN1 and CRWN4 proteins belong to the insoluble nuclear fraction resistant to different concentration of salt and detergent treatment. A similar fractionation method to isolate plant nuclear matrices led to the initial discovery of NMCP1 protein from carrot suspension cells [75]. Also, Ciska *et al.* showed that onion NMCP1 was only observed in the insoluble fraction after a stepwise fractionation of nucleoskeleton preparations in onion tissues, where the extracted nuclei were sequentially treated with detergent (0.5% TritonX-100), a chromatin removal reagent (1M $(NH4)_2SO_4$), and high salt (4M NaCl) [77]. The presence of onion NMCP1 in the residual nucleoskeleton was also demonstrated by the immunogold staining using electron microscopy [77].

Nonetheless, these findings may not be enough to demonstrate the existence of a nuclear matrix system *in vivo*. The primary concern is that the extraction procedures themselves can lead to protein precipitation and form insoluble nuclear residues{Jackson, 1992 #167;Martelli, 1994 #168}. Non-ionic detergents such as Triton X-100 solubilize lipids and may deplete the proteins attached to the outer membrane of the nuclei. However, they do not break nuclear membranes, and Triton treatment had little effect on releasing the nuclear proteins into the supernatant, such

as histone H2B as well as CRWN1 (Figure 4.3A). CHAPS, another detergent frequently used for membrane protein extraction, had little effect in solubilizing CRWN proteins; rather, CHAPS helped preserve CRWN1 and CRWN4 in the insoluble pellet from being degraded during sonication (Figure 4.3B). Finally, high salt treatment itself, which is often used to break the nuclei, had an effect in precipitating the CRWN1 protein (Supplementary Figure 4.1B). This increased insolubility could be due to the biochemical properties of NMCP (CRWN) proteins in reaction to particular extraction conditions. In that case, the apparent rigidity of the NMCP (CRWN)-associated residual framework could be more of an artifact than a true reflection of the nature of the nucleoskeleton *in vivo*. With these caveats in mind, I interpret my extraction experiments results as an indication that Arabidopsis CRWN family proteins are located in the nucleus, and further that their unconventional biochemical properties argue for a potential structural function in nuclei.

My subsequent immunoblot experiments revealed that CRWN4 protein is significantly reduced in *crwn1* mutant backgrounds, indicating that the phenotypes observed in *crwn1* mutants reflect the combined effect of a loss of CRWN1 and a partial loss of CRWN4. Therefore, the observation that *crwn1* induced mis-expressed loci belonging to the pool of genes mis-regulated in the *crwn4* mutant could be due to reduced CRWN4 protein in the *crwn1* mutant, rather than the loss of CRWN1 itself. In this interpretation, CRWN1-like and CRWN4 proteins target similar genomic regions primarily via CRWN4 function. However, several points of evidence support our original hypothesis that CRWN1-like proteins possess independent functions

divergent from the CRWN4 protein, and that CRWN proteins affect transcriptional regulation through a complex set of interactions. First, guard cell nuclear size is reduced in *crwn1* mutants but not in *crwn4* mutants, and CRWN1 plays the major role in controlling nuclear size in leaves [140] . Second, the aggregation phenotype of heterochromatin in the *crwn* nuclei and dwarfism at the whole plant level is exclusively associated with the loss of CRWN1-like function [140] . Third, *crwn1 crwn2* double mutants show more severe whole plant and nuclear morphology phenotypes, as well as more profound mis-expression genome wide, in comparison to *crwn1* mutants (Chapter 3) [78, 140]. This difference is attributable to the addition of the *crwn2* mutation, as our protein immunoblot data did not detect any further reduction of CRWN4 abundance in *crwn1 crwn2* mutants. Therefore, it is likely that CRWN1 and CRWN4 functions in the same complex, collectively regulating the transcription of common loci in Arabidopsis. However, other complexes might also exist which contains either CRWN1 or CRWN4 and are responsible for *crwn1*- or *crwn4*-specific morphological changes.

I used co-immunoprecipitation experiments to demonstrate that CRWN1 and CRWN4 physically interact *in vivo*. Although this approach cannot distinguish between direct and an indirect interactions, the structure of CRWN proteins suggests a simple explanation for a direct interaction via the long coiled-coil domain. This type of oligomerization has been shown in many cases for other proteins containing long coiled-coil domains, such as SMC (Structural Maintenance of Chromosome) proteins, tropomyosins, and lamins. For example, each SMC protein contains a long coiled-coil

motif in the middle for self-folding in an antiparallel manner to form the arm of the V shape dimer [52]. Tropomyosins are a large family of integral components of actin filaments. It consists of rod-shaped coiled-coil hetero- or homo-dimers that lie along the α-helical groove of actin filaments to regulate their function [53, 172]. Lamin proteins are the basic elements of the animal nuclear lamina. *In vitro* experiments showed that lamin A proteins form dimers and polymerize through their long coiled-coil domain. In the case of CRWN proteins, the long coiled-coil region of CRWN1 and CRWN4 monomers could wrap around each other in parallel or anti-parallel fashion, and these heterodimers could form a basic unit to execute *CRWN* function. Alternatively, homo-dimer or polymerized CRWN1 and CRWN4 proteins could interact with each other to form higher-order functional structures. As noted above, it is also possible that other molecules are involved in mediating an indirect interaction between CRWN1 and CRWN4. To address this question thoroughly, a domain analysis of truncated CRWN proteins will be needed to determine which part of the protein is necessary for the interaction. In addition, an immuno-affinity purification coupled with mass spectrometry could be used to examine other molecular components of CRWN1-CRWN4 complexes.

The interaction between CRWN1 and CRWN4 proteins prompted me to examine the mRNA expression level of these *CRWN* genes in different *crwn* mutants. Collectively, the mRNA expression level of *CRWN4* gene is up-regulated in *crwn1* backgrounds, although the amount of CRWN4 protein is significantly reduced in *crwn1* mutants. Another demonstration of the interdependence of CRWN1 and CRWN4 functions is

my observation that *CRWN1* mRNA was reduced to half the normal level in *crwn4* mutant backgrounds. Based on these and other interactions among *CRWN* genes, I propose a model (Figure 4.7) to account for the apparent feedback mechanisms that maintain a balance of CRWN1-like and CRWN4-like functions. The model postulates an equilibrium in complementary CRWN functions via regulation of the quantity of CRWN1-like and CRWN4 mRNA and proteins. In wild type, CRWN1 and CRWN4 play specialized roles; yet both are needed to fulfill the function of CRWN complexes in regulating nuclear morphology and transcription. Our genetic analysis indicates that CRWN2 and CRWN3 possess both CRWN1 and CRWN4 functions, with a bias toward CRWN1-like function. In *crwn1* mutants, where the CRWN1 function is absent, this situation is compensated by a mild up-regulation of *CRWN2* and *CRWN3* mRNA, as well as a reduced amount of CRWN4 protein. Therefore, a balance of CRWN1-like and CRWN4 function is be restored, resulting in relatively few mis-expressed loci in *crwn1* mutants. In *crwn4* mutants, CRWN4 function is lost, with a compensatory reduction in mRNA expression of the *CRWN1* gene. However, this compensation only partially alleviates the excess of CRWN1-like proteins in the nuclei. A similar scenario happens in *crwn1 crwn2* double mutants, where a more complete loss of CRWN1-like function push the ratio between CRWN1-like and CRWN4 functions toward an extreme. A reduction of CRWN4 protein also exist in *crwn1 crwn2* mutants, however, there is no obvious up-regulation of the *CRWN3* gene, indicating that this extreme condition impairs the compensating mechanism. The low quantity and unbalanced pool of CRWN1-like and CRWN4 proteins resulted in dramatic changes of nuclear function, reflected by the mRNA-seq analysis discussed

in Chapter 3. In *crwn1 crwn4* double mutants, however, *CRWN2* and *CRWN3* genes were significantly over-expressed in the absence of both CRWN1 and CRWN4 proteins, suggesting an active compensation to restore the quantity and balance of CRWN1-like and CRWN4 functions. This model is consistent with the morphological and transcription suppression observed in *crwn1 crwn4* double mutants. Further work to test the predictions of this working model will require a more complete understanding of CRWN2 and CRWN3 protein regulation and activity.

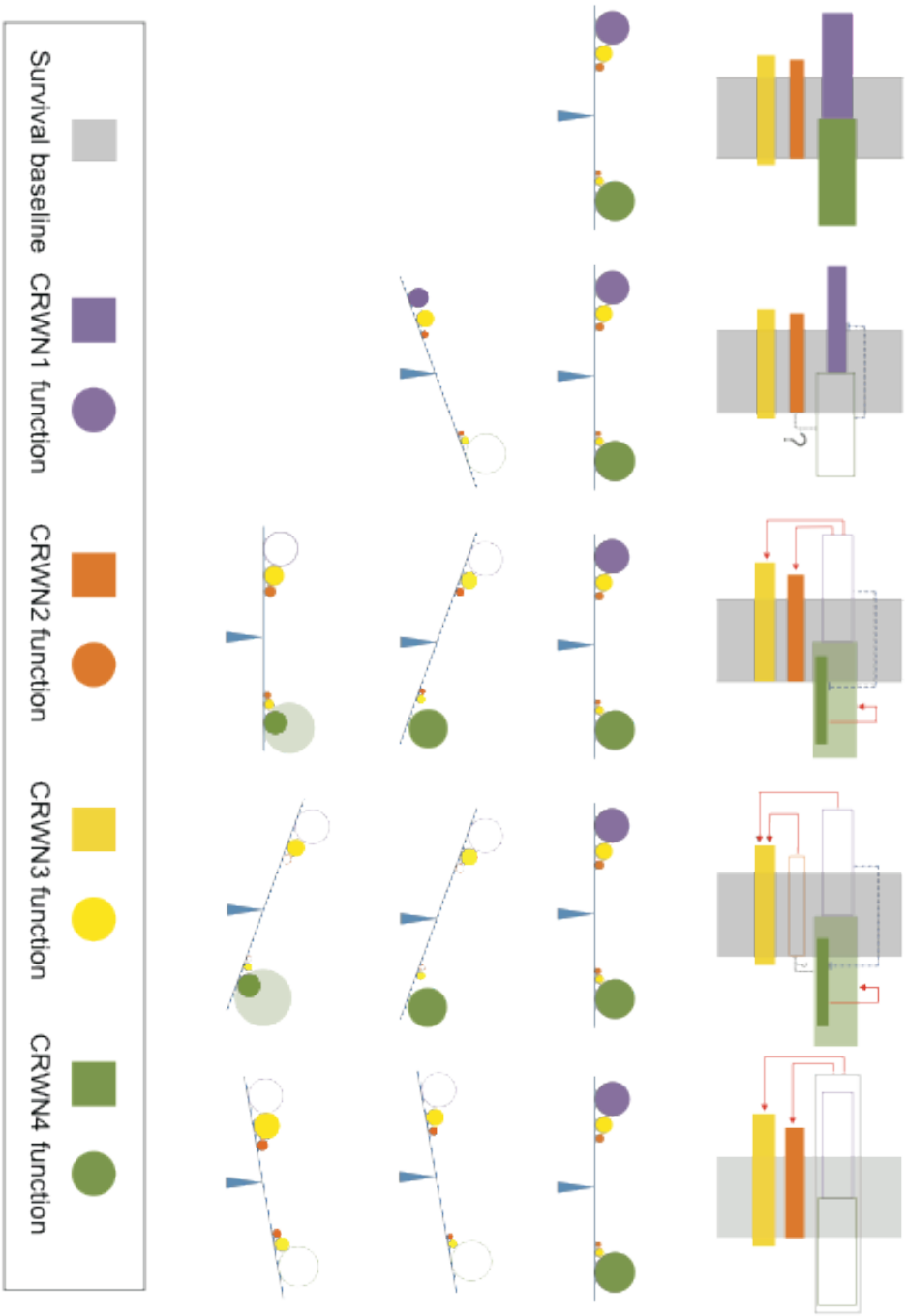


Figure 4.7 A balancing model of CRWN1-like and CRWN4-like functions

Figure 4.7 A series of cartoons describe the genetic interaction (top) as well as the balance of CRWN1-like and CRWN4 functions (bottom) in wild type and representative *crwn* mutants. The rectangle boxes in the models (top) illustrated divergent and redundant functions of *CRWN* genes. The width of the grey area represents the minimum coverage of *CRWN* function that is necessary for the survival of a plant. The size of each colored rectangle reflects the portion of the function that each *CRWN* gene contributes to nuclear organization and transcriptional regulation. Filled rectangles reflect transcription levels of wild type copy *CRWN* genes, while outlined empty rectangles reflect *CRWN* genes with null function in *crwn* mutant backgrounds. In particular, the size of dark green rectangle is reduced in *crwn1* and *crwn1 crwn2* mutant backgrounds against the background of an enlarged light green rectangle, indicating the down-regulation of CRWN4 protein and up-regulation of CRWN4 mRNA when *crwn1* mutation is present. The potential compensatory activation or repression among different CRWN genes is displayed by solid red arrows or dashed blue lines. Other potential regulation is shown by question marks. The cartoons (bottom) illustrate the hypothesized balance of CRWN1-like and CRWN4 functions, depending on CRWN mRNA and protein expression levels in each *crwn* mutant. The loss of balance between *CRWN1*-like and *CRWN4* functions is postulated to lead to extreme morphological and transcriptional changes.

Materials and Methods

Protein sequence analysis

Predicted CRWN proteins sequences of various *Arabidopsis thaliana* ecotypes were downloaded from the publicly available genomic sequences of 856 natural strains on the 1001 genome website (http://signal.salk.edu/atg1001/3.0/gebrowser.php). Predicted CRWN protein sequences of *Arabidopsis lyrata* were obtained from the *A. lyrata* 1.0 genome from the JGI website (http://genome.jgi-psf.org/Araly1/Araly1.home.html). Polymorphic amino acid sites were identified using the multiple protein sequence alignment webserver Clustal Omega [25] (https://www.ebi.ac.uk/Tools/msa/clustalo/).

Plant materials and growth conditions

Seeds from wild type, *crwn1*, and *crwn4* genotypes were planted on MS plates, cold-treated for 3 days in 4°C, and then germinated under long-day lighting conditions (16 h of light / 8 h of dark) at 23˚C in environmental growth chambers.

Nuclear extract preparation

2-week-old seedlings of wild type and *crwn1* and *crwn4* mutants were harvested, and fresh tissue homogenized thoroughly in Honda buffer [173] (0.44 M sucrose,1.25% Ficoll, 2.5% Dextran T40, 20 mM Hepes-KOH pH 7.4,10 mM $MgCl_2$, 0.5% Triton X-100) on ice, and filtered through Calbiochem® Miracloth (EMD Millipore Bioscience) twice. The clear liquid phase was centrifuged at 7000g for 30 minutes in 4°C, and

nuclear pellets were collected and washed three times using Honda buffer until the pellets turned gray. If needed, an additional wash using sucrose gradient nuclear extraction buffer (NEB3 in original protocol) was performed [171].

Antisera development and test

Peptide antigens against the N-terminal regions of CRWN1 and CRWN4 proteins were synthesized by a commercial antisera production company (Proteintech): (CRWN1: MSTPLKVWQRWSTPTKATN; CRWN4: RVLKSPLTEEIMWKRLKD.) These peptides were injected to rabbits to raise the corresponding antisera, which were affinity purified for this study by Proteintech. The nuclear extracts from wild type and *crwn* mutants were boiled in Laemmli buffer for 5 min and resolved by SDS-PAGE gel electrophoresis. The specificity and efficacy of antisera were tested at different concentrations and washing stringency using protein immunoblots (see protocol below).

Salt, detergent and sonication treatment on nuclear extracts

A mock nuclear lysis buffer was made as control (40mM HEPES pH 7.9, 3mM $MgCl_2$, 0.4mM EDTA, 40% Glycerol, add protease inhibitor before use), and each condition was prepared by adding one or a combination of a few of these treatments to the lysis buffer: a particular concentrations of NaCl (0M, 0.5M, 1M, 1.5M, 2M), one or more detergents: Triton X-100 (1%, 2%), CHAPS (1%, 2%, 4%), or SDS (0.05%, 0.2%), a short (5 sec x 10) or long (5 sec x 20) sonication. The extracted nuclei were treated and gently rocked for 30 minutes at 4°C, followed by centrifugation to separate the

supernatant from the insoluble pellet. Proteins were harvested by boiling the supernatants and pellets separately in Laemmli buffer (with 5% beta-mercaptoethanol) for 5 min and resolved on SDS-PAGE gels. Protein immunoblots (see below) were used to determine whether CRWN proteins were solubilized and present in the supernatant samples in each treatment condition.

Protein immunoprecipitation and protein immunoblots

The extracted nuclei were sonicated in immunoprecipitation (IP) buffer (1 mM EDTA, 100 mM Tris-Cl pH 7.4, 10% v/v glycerol, 75 mM NaCl, 0.05% w/v SDS, 0.1% vol/vol Triton X-100), and centrifuged to collect the supernatant. Different CRWN antisera was added into the IP buffer at 100ug/ml and incubated with gentle rotation overnight at 4°C. The next day, to recover the antisera-protein complexes, protein-G Dynabeads® (Life Technologies) were added into each sample with continued gentle rotation for two more hours at 4 °C . Then, the beads were collected using magnetic capture and washed 3 times using IP buffer. Proteins were harvested by boiling the beads in Laemmli buffer [174] (2% SDS, 10% glycerol, 0.01% Bromophenol blue, with 5% beta-mercaptoethanol) for 5 min, and resolved by SDS-PAGE gel electrophoresis. Proteins were semi-dry transferred onto a PVDF (polyvinylidene difluoride) membrane (GE Health Care) and the signals of CRWN proteins were detected using ECL (Enhanced Chemiluminescence) western blotting detection system (GE Health Care).

REFERENCES

1. Dechat T, Pfleghaar K, Sengupta K, Shimi T, Shumaker DK, Solimando L, Goldman RD: **Nuclear lamins: major factors in the structural organization and function of the nucleus and chromatin**. *Genes & development* 2008, **22**(7):832-853.
2. Butin-Israeli V, Adam SA, Goldman AE, Goldman RD: **Nuclear lamin functions and disease**. *Trends in genetics : TIG* 2012, **28**(9):464-471.
3. Rout MP, Field MC: **Isolation and characterization of subnuclear compartments from Trypanosoma brucei. Identification of a major repetitive nuclear lamina component**. *The Journal of biological chemistry* 2001, **276**(41):38261-38271.
4. Batsios P, Peter T, Baumann O, Stick R, Meyer I, Graf R: **A lamin in lower eukaryotes?** *Nucleus* 2012, **3**(3):237-243.
5. Dittmer TA, Stacey NJ, Sugimoto-Shirasu K, Richards EJ: **LITTLE NUCLEI genes affecting nuclear morphology in Arabidopsis thaliana**. *The Plant cell* 2007, **19**(9):2793-2803.
6. Masuda K, Xu ZJ, Takahashi S, Ito A, Ono M, Nomura K, Inoue M: **Peripheral framework of carrot cell nucleus contains a novel protein predicted to exhibit a long alpha-helical domain**. *Experimental cell research* 1997, **232**(1):173-181.
7. Ciska M, Moreno S: **NMCP/LINC proteins: Putative lamin analogs in plants?** *Plant signaling & behavior* 2013, **8**(12).

8. Zhou X, Meier I: **How plants LINC the SUN to KASH**. *Nucleus* 2013, **4**(3):206-215.

9. Graumann K, Evans DE: **Plant SUN domain proteins: components of putative plant LINC complexes?** *Plant signaling & behavior* 2010, **5**(2):154-156.

10. Sakamoto Y, Takagi S: **LITTLE NUCLEI 1 and 4 regulate nuclear morphology in Arabidopsis thaliana**. *Plant & cell physiology* 2013, **54**(4):622-633.

11. Dittmer TA, Richards EJ: **Role of LINC proteins in plant nuclear morphology**. *Plant signaling & behavior* 2008, **3**(7):485-487.

12. Wang H, Dittmer TA, Richards EJ: **Arabidopsis CROWDED NUCLEI (CRWN) proteins are required for nuclear size control and heterochromatin organization**. *BMC plant biology* 2013, **13**(1):200.

13. Ciska M, Masuda K, Moreno Diaz de la Espina S: **Lamin-like analogues in plants: the characterization of NMCP1 in Allium cepa**. *Journal of experimental botany* 2013, **64**(6):1553-1564.

14. Kimura Y, Kuroda C, Masuda K: **Differential nuclear envelope assembly at the end of mitosis in suspension-cultured Apium graveolens cells**. *Chromosoma* 2010, **119**(2):195-204.

15. Tamura K, Fukao Y, Iwamoto M, Haraguchi T, Hara-Nishimura I: **Identification and characterization of nuclear pore complex components in Arabidopsis thaliana**. *The Plant cell* 2010, **22**(12):4084-4097.

16. Burkhard P, Stetefeld J, Strelkov SV: **Coiled coils: a highly versatile protein folding motif**. *Trends in cell biology* 2001, **11**(2):82-88.

17. Harbury PB, Zhang T, Kim PS, Alber T: **A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants**. *Science* 1993, **262**(5138):1401-1407.

18. Schubert V: **SMC proteins and their multiple functions in higher plants**. *Cytogenetic and genome research* 2009, **124**(3-4):202-214.

19. McLachlan AD, Stewart M: **Tropomyosin coiled-coil interactions: evidence for an unstaggered structure**. *Journal of molecular biology* 1975, **98**(2):293-304.

20. Dechat T, Adam SA, Taimen P, Shimi T, Goldman RD: **Nuclear lamins**. *Cold Spring Harbor perspectives in biology* 2010, **2**(11):a000547.

21. Strelkov SV, Schumacher J, Burkhard P, Aebi U, Herrmann H: **Crystal structure of the human lamin A coil 2B dimer: implications for the head-to-tail association of nuclear lamins**. *Journal of molecular biology* 2004, **343**(4):1067-1080.

22. Weigel D, Mott R: **The 1001 genomes project for Arabidopsis thaliana**. *Genome biology* 2009, **10**(5):107.

23. Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT *et al*: **Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines**. *Nature* 2010, **465**(7298):627-631.

24. Hu TT, Pattyn P, Bakker EG, Cao J, Cheng JF, Clark RM, Fahlgren N, Fawcett JA, Grimwood J, Gundlach H *et al*: **The Arabidopsis lyrata genome sequence and the basis of rapid genome size change**. *Nature genetics* 2011, **43**(5):476-481.

25. Jacob E, Unger R: **A tale of two tails: why are terminal residues of proteins exposed?** *Bioinformatics* 2007, **23**(2):e225-230.

26. Pederson T: **Half a century of "the nuclear matrix"**. *Molecular biology of the cell* 2000, **11**(3):799-805.

27. Fiil BK, Qiu JL, Petersen K, Petersen M, Mundy J: **Coimmunoprecipitation (co-IP) of Nuclear Proteins and Chromatin Immunoprecipitation (ChIP) from Arabidopsis**. *CSH protocols* 2008, **2008**:pdb prot5049.

28. Hirano T: **At the heart of the chromosome: SMC proteins in action**. *Nature reviews Molecular cell biology* 2006, **7**(5):311-322.

29. Kalyva A, Schmidtmann A, Geeves MA: **In vitro formation and characterization of the skeletal muscle alpha.beta tropomyosin heterodimers**. *Biochemistry* 2012, **51**(32):6388-6399.

30. Gaudino RJ, Pikaard CS: **Cytokinin induction of RNA polymerase I transcription in Arabidopsis thaliana**. *The Journal of biological chemistry* 1997, **272**(10):6799-6804.

31. Laemmli UK: **Cleavage of structural proteins during the assembly of the head of bacteriophage T4**. *Nature* 1970, **227**(5259):680-685.

32. Sievers F, Higgins DG: **Clustal Omega, accurate alignment of very large numbers of sequences**. *Methods in molecular biology* 2014, **1079**:105-116.

CHAPTER 5

FUTURE DIRECTIONS

In this thesis, I used genetic, molecular and biochemical tools to study the long coiled-coil domain CRWN family proteins in flowering plants *Arabidopsis thaliana*. I showed that different combinations of *crwn* null mutations lead to a series of morphological changes in mutant plants, such as rounded nuclear shape, reduced nuclear area, disturbed endo-reduplication, increased nuclear DNA density, and disrupted heterochromatin organization. My transcriptomic profiling further uncovered various mis-regulated loci, which potentially contribute to these phenotypic alterations.

Phylogenetic analysis revealed the sub-classification of *CRWN1*-like and *CRWN4*-like genes, and this grouping was supported by the divergent morphological changes caused by mutation of *CRWN1*-like genes versus *CRWN4*. Surprisingly, these two clades of CRWN proteins regulate common genomic loci during transcription. At present, I do not know if the transcriptional mis-regulation in *crwn* mutants occurs via direct or indirect mechanisms. In any case, the convergence on common transcriptional target is possibly fostered by the physical interaction between CRWN1 and CRWN4 proteins, which I demonstrated in subsequent biochemical characterization. These data suggested a complementary relationship between

CRWN1-like and CRWN4 functions. I propose a a balancing model of these two complementary functions in organizing the nuclei.

CRWN1 and CRWN4 proteins are both nuclear proteins resistant to high salt and mild detergent extraction, and are candidates for structural components of the nucleoskeleton. To test this hypothesis, more cell biology and biochemistry investigations need to be carried out to understand the structural status of CRWN proteins *in vivo*, and the identity of working partners for CRWN proteins. FRAP (Fluorescence Recovery After Photobleaching), immuno-gold localization of CRWN proteins using EM (Electronic Microscopy), and more sophisticated cell fractionation could help illustrate the physical properties of CRWN proteins in the nuclei. Immuno-affinity purification and mass spectrometry techniques could be used to identify interacting partners in CRWN complexes, as well as reveal possible post-translational modifications. A candidate approach by immunoprecipitation could also be taken to explore CRWN-interacting proteins, such as SMC subunits, CRWN2, CRWN3, SUN1, and SUN2 proteins. Alternative systems could be used for validation, such as yeast-two-hybrid studies, split GFP or luciferase complementation in tobacco, and *in vitro* pull down of candidate proteins expressed in bacteria.

Another important aspect of CRWN function relevant to transcriptional regulation involves chromosome organization. Chromosome immunoprecipitation (ChIP) of CRWN1 and CRWN4 proteins could help understand whether any CRWN complexes interact with chromatin directly, and could determine whether CRWN complexes

target specific genomic regions. In addition, whole genome bisulfite sequencing (WGBS) of *crwn* mutants, as well as profiling of various histone modifications, would also provide information about the potential interaction between CRWN proteins and the epigenome. It would also be interesting to compare interacting loci from these profiling results with genomic regions recognized by other proteins (*e.g.*, SMC subunits). Any shared profiles would aid in identification of partner proteins functionally interacting with CRWN, and help dissect the mechanisms through which CRWN proteins regulate nuclear organization.

Other interesting directions to explore include determining whether the physical properties of *crwn* nuclei differ from wild type nuclei, and if so, whether these changes are primary or secondary, and are associated with alteration in nuclear function. A second avenue to pursue is the long-term consequences of crwn-mediated nuclear changes. For instance, do epigenetic and genetic variation accumulate in *crwn* mutant backgrounds due to the mis-expressed epigenetic modifiers and DNA replication machinery. A third set of possible experiments include study of genomic level reorganization in *crwn* mutants. To understand whether *crwn* mutations alter the organization of chromosomes, chromatin conformation capture (3C) and related techniques could be utilized to study the chromosomal interactome in *crwn* mutant backgrounds.